

THIS WEEK

EDITORIALS

REPUTATIONS The dos and don'ts of online profile management **p.124**

HOW MANY? The dubious assumptions behind population estimates **p.125**

DESERT SPRING New species emerges from whiptail breeding **p.126**



Those who can

An initiative from Scientific American aims to find 1,000 scientists to visit schools, help teachers and boost US education.

Many developed nations face chronic problems in high-school science education. Researchers in the United States, in particular, will be only too aware of the nation's sliding student scores in science, technology, engineering and mathematics (STEM) education. What may surprise, however, is that a truly powerful resource is available that, as one science and maths teacher exulted on Twitter, "will be a HUGE gamechanger for the good".

You.

How can you hope to inspire kids to love science when so many educators and policy leaders have struggled? And how can you possibly spare large chunks of precious time from research and grant writing to do so?

Your experience is invaluable, and you will have help. Just as large research projects involve many hands, so an initiative from the magazine *Scientific American* called '1,000 Scientists in 1,000 Days' seeks to recruit an army of researchers to help in targeted, concrete ways. The initiative is part of the 'Change the Equation' programme, which was set up in part to realize President Barack Obama's campaign mission to boost private and philanthropic participation in STEM education. Think of it as a kind of science corps to support the growth of developing minds. A sign-up form can be found at www.scientificamerican.com under the education tab. *Scientific American* is a sister publication of *Nature* within the Nature Publishing Group. We at *Nature* are glad to promote this initiative.

What can scientist volunteers do? Perhaps they could spend an hour in a local classroom or school auditorium talking about a typical day in the lab — thereby helping to demystify the world of science for children. They could give a local school board advice about curricula or specific research areas. They might simply answer questions by e-mail, teleconference or Skype. How scientists participate, and how frequently, will be up to them. By the beginning of the new school year, around September, *Scientific American* will be able to connect educators with experts.

Perhaps some additional background would help to convince you that it is worth the time to visit classrooms. Last year, a report by the US National Academies found that the United States ranks 27th of 29 wealthy countries in the proportion of university students who graduate with degrees in science or engineering. It called on federal and state governments to improve teaching in maths and science by targeting early-childhood education and the public-school curricula, and by supporting teacher training in crucial subjects.

In the younger grades, many US science teachers have no science training: in 2004, only 40% of fifth- and 80% of eighth-grade students were taught maths and science by teachers with a degree or certificate in their teaching field, according to the most recent figures from the National Science Foundation.

What is more, teachers have to juggle the often-conflicting demands to 'teach to the test', which requires a lot of learning by rote, with the need to imbue students with the inspiring wonder of science — and

the process-driven critical thinking and evidence collection that proper research requires. Educators also wrestle with anti-science demands to 'teach the controversy' in disciplines such as evolution and climate change. According to the National Center for Science Education, at least eight anti-evolution bills have been introduced in US state legislatures since the beginning of 2011.

Obama has said that "winning the future" for an beleaguered US economy will require investment in research, innovation and education (see *Nature* **470**, 313–315; 2011). In his January State of the Union address, he said that science-fair winners should be as celebrated as Super Bowl champions, and he has added hands-on science activities to the Easter Egg Roll, an annual event for children on the White House lawn since 1878.

Announcing the Change the Equation programme last September, Obama said: "Our success as a nation depends on strengthening America's role as the world's engine of discovery and innovation." The 1,000 Scientists in 1,000 Days programme is part of a broader initiative from Nature Publishing Group, called Bridge to Science, which includes five other schemes collectively focused on addressing the needs of parents, educators, policy leaders and, ultimately, the collective progress of science itself.

Although the programme initially targets US needs, the sign-up form is open to scientists anywhere, and the scheme could expand further. More than 230 scientists have already signed up. We welcome their generosity and interest, and hope you will join them. ■

"How scientists participate, and how frequently, will be up to them."

Value judgements

The scientific endeavour needs to deliver public value, not just research papers.

The concept of a return on research investment has acquired a sharper edge since the global financial slump began. But an assessment of those returns should include more than knowledge for its own sake and economic growth — as highlighted by a timely series of articles in the latest issue of the journal *Minerva*.

Under the intellectual and editorial leadership of the policy scientists Barry Bozeman of the University of Georgia in Athens and Daniel Sarewitz of Arizona State University in Tempe, the journal presents case studies that analyse a broader way to measure returns on investment: public values. These public values include not only the commonly discussed knowledge and economic criteria, but also

information useful to decision-makers, participation in agenda-setting by stakeholders and communication to the public in general. In fields as diverse as hurricane research and nanomedicine, the contributors demonstrate how initial ambitions to target these values get lost as projects unfold. Too often, the pull of the heavyweight 'science for its own sake' and economic agendas crushes wider intentions.

Policy-makers, funders and scientists should take note. For example, a paper by Ryan Meyer, also a policy scientist at Arizona State University, focuses on the failure of the US government's Global Change Research Program to deliver broad public value (*Minerva* 49, 47–70; 2011). Basing his studies on public statements and private interviews with researchers and political decision-makers, Meyer says that US climate programmes have in the past two decades benefited from public investment of more than US\$30 billion, but have largely failed to produce information and participation in the forms that policy-makers and the public wanted. The notion that society considers any advance in knowledge to be inherently good — even if the science fails to meet the objectives and priorities it was meant to address — cannot be sustained, says Meyer.

Coincidentally, at a workshop last week in Hamburg, Germany, a gathering of climate scientists, policy experts and philosophers of 'post-normal science' articulated a similar perspective. Science becomes 'post-normal' when facts are uncertain, stakes high, values in dispute and decisions urgent; in such cases, societal needs must be taken into account to avoid costly mistakes. Climate research certainly fulfils this definition. But, according to the workshop participants, most climate researchers continue to act as if purely scientific values are, and will always be, adequate to set the agenda.

Instead, researchers and, especially, their funders must embrace the idea that public and stakeholder participation can help to define research priorities. And they must do more to track and communicate all outcomes. Policy-makers need to ensure that those with direct needs for climate-related information — businesses, regional planners, government departments — have a greater say in the kind of services and knowledge that they expect publicly funded researchers to

produce, and in assuring the quality and relevance of what is delivered.

The agenda here is to broaden rather than restrict climate research: to ensure that, alongside studies driven by the priorities of creative scientific imaginations, research also enhances public confidence in society's use of science. A fuller consideration of these issues will also maximize science's public value, in the form of successful collaboration across

disciplinary boundaries, delivery of information useful to stakeholders, and transparency and well targeted communication.

Grant-giving agencies need to take a fresh look at how to address these issues. Lay representation in decision-making, as has already been tried on a limited scale at the US National Institutes of Health, is one way to increase public involvement. Previous exercises in engagement on nanotechnology have

led to positive, albeit restricted, outcomes (see *Nature* 448, 1–2; 2007).

But the problem goes deeper than public engagement. The Global Change Research Program has involved stakeholders at various stages, but, says Meyer, "it is not clear that the US currently has the institutional capacity to achieve the public values promised about climate science". If that is indeed the issue then, given the sheer scale and breadth of US capacity in these topics, one has to wonder whether anyone will be able to achieve what these authors are seeking.

As is often the case, policy researchers are good at identifying problems but slower to reveal how to improve matters. To his credit, Meyer offers some practical ways for the programme to improve the leverage of, and incentives for, a broader participation in delivering public value.

More importantly, these studies highlight a significant deficit in current typical appraisals of science and technology outcomes. They should serve as cautionary tales about the danger of scientists' interests, deliberately or otherwise, becoming too dominant in determining outcomes. And they introduce ways to assess failures in social returns on investment that, one can only hope, will help to improve science's public value. ■

polishing. A survey for *Nature* this week suggests that up to 10% of scientists have considered using external services to manage their online reputations (see page 138). Others have edited their own biographies on the online resource Wikipedia — a practice that is frowned upon — or inserted references to their own work.

Online reputation is important to most researchers, and about 10% of respondents to our survey complained that they or their work have been misrepresented on the Internet. The web has a long memory, and rumours, lies and bad information can spiral out of control to be remembered by posterity.

Through responsible use of blogs and social media, researchers have the power to chip away at misperceptions. This isn't about flame wars or trolling comment boards — a lifetime could be spent telling people on the Internet that they're wrong — but rather involves getting the right facts out there, and citing and linking to the best, most trustworthy sources of information.

Such diligence can also benefit scientists as members of a professional community. Researchers who make sure that personal and institutional websites, blogs and social-media pages are accurate and honest will enhance the usefulness of web searches by pushing the most relevant and trusted information to the top. This can make it easier for scientists to find one another for collaboration and reviewing papers, and to locate and fill jobs.

Enhancing visibility and promoting a digital image may strike some as unsavoury, but it is not. Researchers are right to promote themselves and their work in a reasonable capacity. The Internet has provided a tremendous tool to do this effectively. And a little more besides. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq

Online image

The Internet offers ways for researchers to steer public perceptions, for bad and good.

"My reputation grows with every failure," said Irish dramatist George Bernard Shaw. And that was before the Internet. Shaw would no doubt be amazed by how quickly electronic bulletin boards such as Facebook and Twitter can now spread the word of deeds both good and bad.

Take the example of Anil Potti, a cancer specialist who resigned his post at Duke University in Durham, North Carolina, last year after it was revealed that he had lied on his CV about being a Rhodes Scholar, and whose research, which had been the basis for clinical cancer trials, was called into question. The Internet should be an unforgiving place for Potti; yet instead of damning news reports, a search on his name throws up a series of plaudits. Numerous websites have been set up to salute his conclusion that smoking causes lung cancer, and to broadcast the information that he enjoys spending time with his family and finished top of his class in high school. Prospective employers who check his background online can still find details of his resignation — but will have to look a little bit more carefully.

Job done, then, for Online Reputation Manager, the company listed as an administrative contact in the registration details for two of the more positive sites. Potti is surely grateful for their efforts.

He is not alone in benefiting from a little electronic-profile



Dubious assumptions prime population bomb

The United Nations says there could be 10 billion people on Earth by the end of the century. Fred Pearce finds problems in its analysis.

The latest global population projections, published by the United Nations last week, say that the world will be awash with 10.1 billion people by 2100, a billion more than previously supposed. Already, there is talk again of a ticking population time bomb.

But a closer look at the assumptions behind this scenario shows it to be perverse and contradictory. In fact, it looks more like a political construct than a scientific analysis.

The heart of the problem is this: the new UN estimates record that both world population and global fertility rates are currently slightly lower than presumed when the last projections were made two years ago. Yet, they project significantly higher growth rates than those estimated two years ago.

This paradox is created by a seemingly arbitrary change in assumptions about future fertility that requires a proper explanation. And quickly.

Plans to cope with an increasing array of global challenges — not least climate change and food policy — are predicated on the UN's demographic projections. The past few years have seen a plethora of scientific papers asking 'can the world feed 9 billion?' It won't be long before the work is revisited to see whether we can feed 10 billion.

We are doing quite well at defusing the population bomb. Women today, on average, have half as many babies as their grandmothers did. World fertility has fallen from 4.9 children per woman in the early 1960s to an expected 2.45 between 2010 and 2015, a projection revised down from the 2.49 figure of two years ago.

The trend is near-universal. With childhood diseases such as measles and tetanus in retreat, for the first time in history most children get to grow up. Population quadrupled in the past century as this happened. But now women are learning to adjust to falling infant mortality and having fewer children. Other factors include urbanization. On a peasant farm in Africa, young children are an economic asset, minding the goats or fetching and carrying. Once families move to the cities, children are a liability, requiring years of education to get a job. Fertility rates are much lower in cities.

Falling fertility doesn't instantly translate into fewer babies. That is because of the huge demographic bulge of twentieth-century baby boomers — now adult and fertile. But as they age, and if fertility rates continue to fall, population growth must subside and could go into decline.

The key questions are how fast and how far fertility will fall. As the UN notes, "small variations in fertility can produce major differences in the size of populations over the long run". That is why the assumptions built into the new projections are so crucial.

The UN's previous 'medium variant' projection, published in 2008, concluded that world population would rise from the present 7 billion and peak in mid-century at around the 9.15 billion

expected in 2050. The new projection finds no peak. Instead, world population reaches 9.3 billion in 2050 and 10.1 billion in 2100, with further growth still in the works.

The UN has yet to publish its detailed reasoning, but a collection of frequently asked questions issued alongside the new projections says that most of the difference is due to an upward revision of its fertility forecasts — a revision unrelated to current trends.

There is history to this. For many years, demographers reckoned that world fertility was headed inexorably for the rich-world replacement level of about 2.1 children per woman. But in the past 30 years, this has looked increasingly like too high a number. In almost all developed countries, fertility rates have fallen to well below replacement levels. Despite a minor bounce-back in recent years, most of Europe remains below 1.5.

With much of Asia and Latin America on the same path, almost a decade ago the UN rethought the 2.1 end point. In 2003, its UN population division, under then-director Joseph Chamie, decided that its 'medium variant' projection should instead assume convergence at 1.85. It was a compromise, Chamie told me. Some argued for 1.6, whereas others wanted to retain 2.1. The latter group, he said, feared that a low estimate would send the 'wrong message' that our population worries were over.

The projections made in 2008 retained the figure of 1.85, but it has now reverted to 2.1 — the predominant reason for the leap from 9 billion to 10 billion. The assumption now is that countries with higher fertility rates will fall to

the 2.1 figure and not below, while those below will rise to reach it.

Is this realistic? As Joel Cohen, a demographer at Columbia University in New York, put it in 2002: "No case is yet known of a population with fertility above replacement level that converged to replacement level and then stayed there." That remains the case. Chamie this week said he had seen "no compelling evidence" to justify a return to the 2.1 figure.

The UN boasts that its new projections have incorporated a more probabilistic approach into the model. That is good. But, as the UN makes clear, the model "incorporated the additional assumption that, over the long run, replacement-level fertility would be reached". In other words, the crucial new fertility end point of 2.1 did not emerge from the new probabilistic analysis. It was imposed on it, and the UN should explain why. ■

**IN ALMOST ALL
DEVELOPED
COUNTRIES, FERTILITY
RATES HAVE FALLEN
TO WELL BELOW
REPLACEMENT
LEVELS.**

GO NATURE.COM
Discuss this article
online at:
go.nature.com/c6jt9k

Fred Pearce, environment consultant for New Scientist, is author of *Peoplequake* (Eden Project Books, 2010) and is speaking at a conference on population footprints at University College London on 25–26 May (www.populationfootprints.org).
e-mail: pearcefred@hotmail.com

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

SPECTROSCOPY

NMR without the magnet

To obtain highly sensitive information on molecular structure and function, nuclear magnetic resonance (NMR) relies on a strong magnetic field. However, Alexander Pines at the University of California, Berkeley, and his co-workers have achieved high resolution and intense NMR spectra at zero magnetic field.

In conventional NMR, the external field partially aligns the nuclear spins in the molecule of interest. The authors' technique uses a polarizing agent to do roughly the same thing. The agent, parahydrogen, reacts with the molecule, inducing a greater degree of polarization of nuclear spins than conventional NMR, thereby enhancing the signal. The signals are then detected with a sensitive atomic magnetometer.

Using their technique, the authors were able to distinguish the structural features of several compounds. They say their method could be used as a low-cost, portable alternative to standard NMR machines.

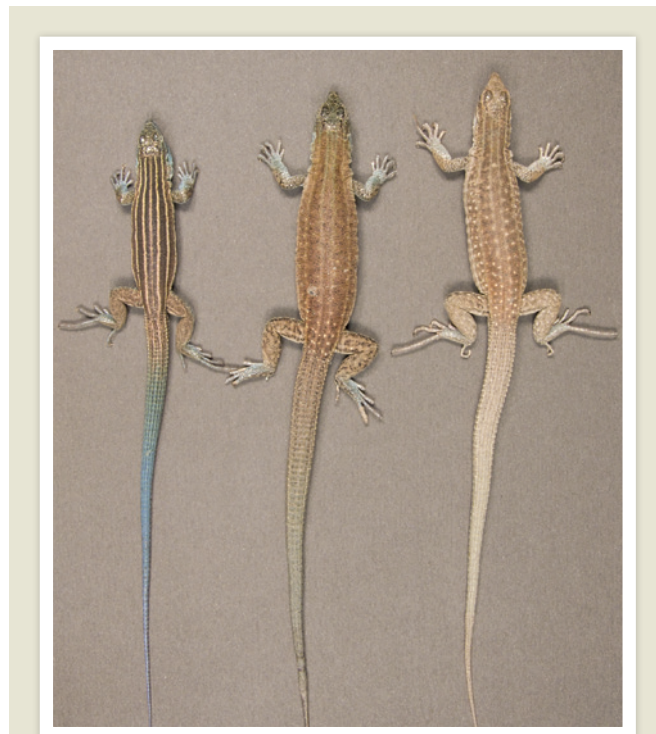
Nature Phys. doi:10.1038/nphys1986 (2011)

ATMOSPHERIC SCIENCE

Methane rises from wetlands

Natural tropical wetlands were the main source of a marked rise in global atmospheric methane levels observed during the past few years.

After a decade of stability, atmospheric concentrations of the potent greenhouse gas began to climb in 2007. Philippe Bousquet at the French National Centre for Scientific Research in Gif-sur-Yvette, France, and his group



EVOLUTION

To make a new species

By mating a female lizard that has three sets of chromosomes with a male that has just the typical two, researchers have created a new species.

Several species of all-female whiptail lizard roam the deserts of New Mexico. They typically reproduce through an asexual process called parthenogenesis. But matings between all-female species and males of other species have, in the past, created 'triploid' hybrids — which have three sets of chromosomes — that then produced sterile offspring. Peter Baumann and his team at the Stowers Institute for Medical Research in Kansas City, Missouri, mated one of these triploid lizards (*Aspidoscelis exsanguis*; pictured right) with a male *A. inornata* lizard (left) to create a new species of whiptail (centre) that has four sets of chromosomes.

The tetraploid lizards went on to produce healthy, genetically identical daughters asexually. Surprisingly, the daughters could also reproduce in this way. The results may help to explain how species with more than two sets of chromosomes evolve.

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.1102811108 (2011)

used atmospheric chemistry and transport models, along with an ecosystem model of wetland methane emissions, to infer the contributions of natural and anthropogenic

sources to the rise.

They found that, in 2007, emissions from tropical wetlands accounted for two-thirds of the 21-million-tonne increase in methane

over the 1999–2006 average. And, in 2008, the wetlands contributed about 50% of the 18-billion-tonne excess. The authors suggest that growth in wetland area due to increased rain is driving these elevated emissions.

Atmos. Chem. Phys. 11, 3689–3700 (2011)

GENOMICS

Gene for malarial drug dodging

A genome-wide study of malaria parasites has revealed a novel gene that confers drug resistance.

Pardis Sabeti at the Broad Institute in Cambridge, Massachusetts, and her team analysed 50 *Plasmodium falciparum* parasites from three continents and found 11 genes associated with resistance to antimalarial drugs. When they transferred one of these genes — which had not previously been associated with resistance — into parasites susceptible to drugs, the parasites became more resistant.

The protein encoded by the gene is located in the parasite's surface membrane, although its function is unknown. However, it does not seem to be a transporter protein such as those already known to pump drugs out of cells.

PLoS Genet. 7, e1001383 (2011)

ORGANISMAL BIOLOGY

Sea urchins 'see' with their feet

Sea urchins can respond to light despite having no obvious eye structure. Researchers have now pinpointed a possible mechanism for this, involving light-sensitive cells in the creatures' numerous tube feet.

Maria Arnone at the Anton Dohrn Zoological Station in Naples, Italy, and

W. B. NEAVES

M. L. ARNONE



her colleagues show that the feet of *Strongylocentrotus purpuratus* (pictured) contain a type of photoreceptor cell found in many other animals, including vertebrates. The researchers found that two key photoreceptor genes (pictured in dark purple and red) are expressed in the feet.

The photoreceptor cells connect to the animals' nervous system (green). The authors suggest that the entire sea urchin can function as a large compound eye, and that the shadow cast by the animal's opaque skeleton on some of the light-sensing cells may enable directional vision. *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.1018495108 (2011)

PHOTONICS

Light switch at the speed of light

Optical devices promise to be faster than their electronic counterparts, but to achieve this, swifter switches are needed to route the flow of information encoded by light. Georgios Cstis and Willem Vos at the University of Twente in the Netherlands and their colleagues have built a switch that changes from off to on to off in just one-trillionth of a second. They used a laser beam to change the refractive index of a microcavity made from gallium arsenide and aluminium arsenide layers, thereby switching the wavelength of light the cavity transmits most readily.

The device is limited by the speed of light in the cavity rather than by the response time of the switch's materials.

On the basis of the results, photonic chips could switch 100 times faster than state-of-the-art electronic chips.

Appl. Phys. Lett. 98, 161114 (2011)

CELL BIOLOGY

Deeper insight into a single cell

By using a variety of transition metals to tag cellular proteins, researchers have simultaneously measured up to 34 characteristics in individual cells. That's more than double the number tracked with previous methods.

Flow cytometry has long been an important tool for discovering and characterizing subtypes of cell. It uses fluorescent markers to concurrently follow the activities of up to 15 cellular parameters. In 'mass cytometry', however, Garry Nolan of Stanford University in California and his colleagues used mass spectrometry to measure changes in the levels of metal-tagged proteins. Mass spectrometry has high resolution, allowing many signals to be measured at once.

The authors monitored fluctuations in 34 parameters in healthy human bone-marrow cells, and uncovered several new cell-signalling events. They also followed the cells' responses to the cancer drug dasatinib.

Science 332, 687–696 (2011)

MOLECULAR BIOLOGY

Protein maker and gene regulator

Long thought to simply assemble cellular proteins from RNA molecules, ribosomes may also control the expression of key development genes.

Maria Barna at the University of California, San Francisco, and her team discovered that mouse strains with short, kinked tails (pictured) and an extra rib are deficient in a protein called RPL38 — one of the dozens

COMMUNITY CHOICE

The most viewed papers in science

NANOTECHNOLOGY

Solar cells improve with acid



A sprinkling of nitric acid doubles the efficiency of hybrid silicon–nanotube photovoltaic cells.

Typical hybrids, made of networks of carbon nanotubes laid on crystalline silicon, might be cheaper to make than conventional all-silicon cells. Yet, at best, they convert a meagre 6–7% of sunlight's power into electricity, well under half the performance of commercial silicon cells. Anyuan Cao at Peking University, Beijing, Dehai Wu at Tsinghua University, also in Beijing, and their colleagues show that adding dilute nitric acid boosts the hybrids' efficiencies to 11–13%. The acid improves contact between the silicon and the nanotubes, providing more paths for charge carriers to exit the cell and produce external current.

Nano Lett. doi:10.1021/nl2002632 (2011)

that make up the ribosome. The defect seems to limit the production of some Hox proteins, which help to form the body plan. In normal mouse embryos, RPL38 is expressed at its highest levels in tissues in which the short-tailed mutants show defects, such as the vertebrae. Moreover, the expression of many of the ribosome's 79 constituent proteins varied widely from tissue to tissue in these embryos, suggesting that ribosomes may have specialized functions.

Cell 145, 383–397 (2011)

(corn) production, for example, is estimated to be 3.8% lower than it would have been if Earth were not warming.

David Lobell at Stanford University in California and his co-authors analysed links between national yields and temperature and precipitation trends from 1980 to 2008. They estimate that, despite the fertilizing effect of increased carbon dioxide in the atmosphere, wheat production has dropped by 2.5%. However, thanks to the CO₂ boost, yields of rice are up by 2.9% and soya beans by 1.3%. The United States, which produces about 40% of the world's soya and maize, has so far been unaffected because its crop-growing regions haven't warmed much in summer.

Overall, the changes wrought by the warming climate during the study period bumped up food commodity prices worldwide by about 6.4%, the authors suggest.

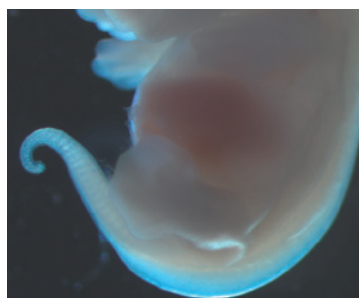
Science doi:10.1126/science.1204531 (2011)

For a longer story on this research, see go.nature.com/rk1fni

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch



AGRICULTURE

Heat hurts crop production

Climate change has suppressed food production around the world during the past three decades, according to modelling work. Global maize

M. BARNA

SEVEN DAYS

The news in brief

POLICY

Nuclear shutdown

Obedying a 6 May request from Prime Minister Naoto Kan, Japanese utility Chubu Electric Power has suspended operations at its Hamaoka nuclear power station, which sits in an area considered overdue for a large earthquake. Three reactors at the plant, which some seismologists have dubbed the “most dangerous” in the country, will not be restarted until Chubu completes existing plans to upgrade sea-wall defences there. The suspension has worried both utilities and local industry, who fear for their electricity supply if other nuclear plants under inspection following the Fukushima disaster are also ordered to close.

Spanish shake-up

A wide-ranging bill that updates 1986 legislation on Spain's science system was expected to pass Congress this week, after the Senate unanimously approved it on 4 May. The bill — a pet project of science minister Cristina Garmendia — would create a state research agency with the power to grant independent funding and allow scientists to move more easily between private and public sectors. After petitions from thousands of researchers, it will also ensure that all scientists, including doctoral students, are hired through contracts rather than fellowships. See go.nature.com/iepb8n for more.

Clinic shut down

One of the world's most notorious stem-cell therapy centres had to cease operations last week, after German law caught up with its widely condemned practices. The XCell-Center, which has units in Düsseldorf and Cologne,



G. GORDON/AFP/GETTY IMAGES

Pacts to curb illegal logging

Liberia's logging industry was once used to fund bloody civil wars that killed hundreds of thousands of people. But this week the country became only the fifth from Africa to sign a trade agreement with the European Union to stop illegal timber exports. Last week, Indonesia became the first Asian country to do the same. The deals are meant to ensure that timber

reaching Europe is sourced legally; measures include electronically tagging trees (pictured). But none of the six countries that have signed the pacts has yet started producing licensed timber. And much of Liberia's wood ends up in China, which does not ask for licences, yet forwards products made from the wood to Europe. See go.nature.com/xwkhck for more.

injected stem cells from bone marrow into the brain, spinal cord and other body parts of patients. An 18-month-old boy died after treatment at the Düsseldorf centre last year. Despite numerous media exposés, the centre's activities continued; but it has now run out of time to apply for a special licence for tissue engineering, under 2007 European regulations implemented in German law in 2009. See go.nature.com/uku3t5 for more.

Australian budget

The feared Aus\$400-million (US\$430-million) cuts in government funding for medical research have not appeared in Australia's 10 May federal budget, to the relief of thousands of scientists who

had rallied in April to protest against the rumoured cuts. But as *Nature* went to press, it seemed that the four-year budget — which aimed to cut the national deficit — did not increase research spending either. The Australian and New Zealand bid to host the Square Kilometre Array radio telescope received US\$40.2 million, in one notable initiative. See go.nature.com/9bqrb2 for more details.

Renewables report

The Intergovernmental Panel on Climate Change has released its first major report since 2007. The 9 May study examined the future of six renewable energy sources (biomass, hydro, wind, solar, geothermal and oceans). More than half of the

164 future scenarios presented in the report suggest that these sources will provide more than 27% of the global energy supply by 2050. In the most optimistic forecast, the proportion rises to 77%. See page 134 for more.

Integrity deadline

After a slow start, the White House Office of Science and Technology Policy (OSTP) is suddenly demanding fast progress on US President Barack Obama's scientific integrity initiative, launched in March 2009. John Holdren, the OSTP's director, said on 5 May that agencies have 90 days to submit draft policies on scientific integrity. So far, 6 out of the 31 agencies affected reportedly have such policies, and only one, the Department

INTEL of the Interior, has a final policy. See go.nature.com/s547fr for more.

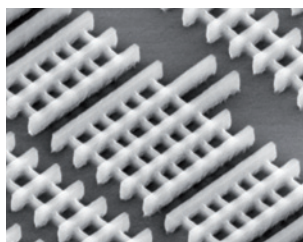
Tevatron shutdown

The Tevatron proton–antiproton collider at Fermilab near Batavia, Illinois, will continue to run until 30 September, when the US fiscal year ends, said Pier Oddone, director of the lab, on 5 May. The US Department of Energy had said in January that it would not fund the collider into 2012, but even a full 2011 running schedule could not be confirmed until Fermilab received its budget breakdown from the April deal made by legislators in Congress.

BUSINESS

Stem-cell trials

California's state stem-cell agency can for the first time say that it is funding a clinical trial. On 4 May, the board of the California Institute for Regenerative Medicine (CIRM) in San Francisco voted to give a US\$25-million loan to Geron of Menlo Park, California, which in 2009 was the first company to get US approval to undertake a clinical trial involving human embryonic stem cells. When the creation of CIRM was put to a public vote in 2004, translation of stem cells from research tools to therapies was a major selling point.



3D transistors

Computer-chip manufacturer Intel has announced that it will mass-produce three-dimensional transistors for its next generation of chips. Usually, the channel along which electric current passes in a transistor is flat; in the 'Tri-Gate' transistors (pictured), it protrudes from the surface so that the 'gate' that switches current on or off wraps around the channel on three sides. The concept — developed in the late 1990s by Chenming Hu at the University of California, Berkeley, and his colleagues — should allow smaller components to be packed more closely on chips. See go.nature.com/2omjfp for more.

RESEARCH

Cholera in Haiti

The cholera epidemic currently raging through Haiti was inadvertently introduced to the country through faecal contamination of river water, a four-member

panel appointed by the United Nations concluded in a report published on 4 May. The report pointed to probable leakage from latrines at a riverside United Nations peacekeepers' camp. However, it stopped short of directly accusing Nepalese soldiers in the camp, who are widely suspected of carrying in the strain (which matches cholera strains circulating in Nepal). The outbreak — the first in Haiti in nearly a century — had by mid-April killed almost 4,900 people and made 286,000 ill.

Gravity probe B

NASA announced on 4 May that its Gravity Probe B mission — conceived and funded five decades ago — had at last confirmed predictions of general relativity. The US\$750-million satellite flew from 2004 to 2005, but it took researchers six years to unpick systematic errors from the data it had collected. Meanwhile, other experiments had already matched its precision. See page 131 for more.

US veterans biobank

The US Department of Veterans Affairs (VA) hopes to enrol 1 million military veterans to form what could be the world's largest research database linking genetic profiles with health records. On 5 May it announced

COMING UP

14–18 MAY

The science and politics of protecting marine life is the focus of the 2nd International Marine Conservation Congress in Victoria, Canada. go.nature.com/ukzje9

16–24 MAY

The 64th World Health Assembly, meeting in Geneva, Switzerland, includes a meeting to decide whether to destroy smallpox stocks. go.nature.com/th3feu

16–19 MAY

Eighteen Nobel laureates are participating in a symposium on global sustainability in Stockholm. go.nature.com/f7mow7

that the Million Veterans Program, already launched at the Boston VA Medical Center, would be expanded across the nation over the next 5–7 years. If it reaches its goals, the programme will be larger than other major personalized medicine initiatives, such as the UK Biobank, which has 500,000 enrolled volunteers.

PEOPLE

New to US academy

Among 72 members elected to the US National Academy of Sciences on 3 May were geneticist George Church of Harvard Medical School in Boston, Massachusetts, and Neil Shubin, an evolutionary biologist at the University of Chicago in Illinois. Stem-cell pioneer Shinya Yamanaka at the University of Kyoto in Japan was among 18 new foreign associate members. See go.nature.com/ax7qth for more.

► **NATURE.COM**

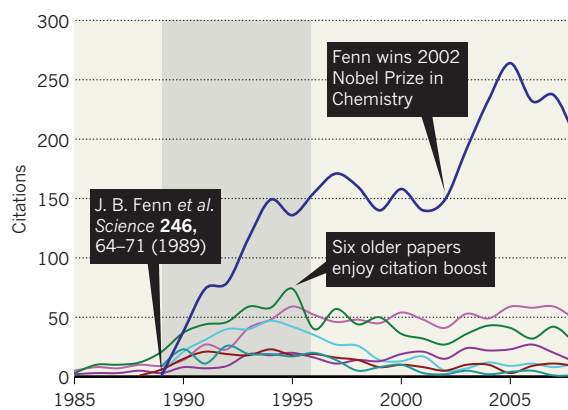
For daily news updates see: www.nature.com/news

TREND WATCH

A pivotal paper by chemist John Fenn saw huge numbers of citations almost immediately after its 1989 publication, but citations of at least six of the future Nobel laureate's older papers also rose (see chart). The same attention boost, or 'citation cascade', is evident in records of more than 100 other Nobel prizewinners, says a study published on 4 May (A. Mazloumian *et al.* *PLoS ONE* 6, e18975; 2011). The finding may shed light on how scientific reputations emerge. See go.nature.com/gphgcp for more.

HOW SCIENTIFIC PRESTIGE EMERGES

A landmark paper quickly boosts an author's citations — but it also generates a cascade of citations and attention for older papers.

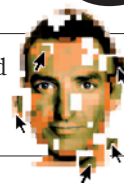


NEWS IN FOCUS

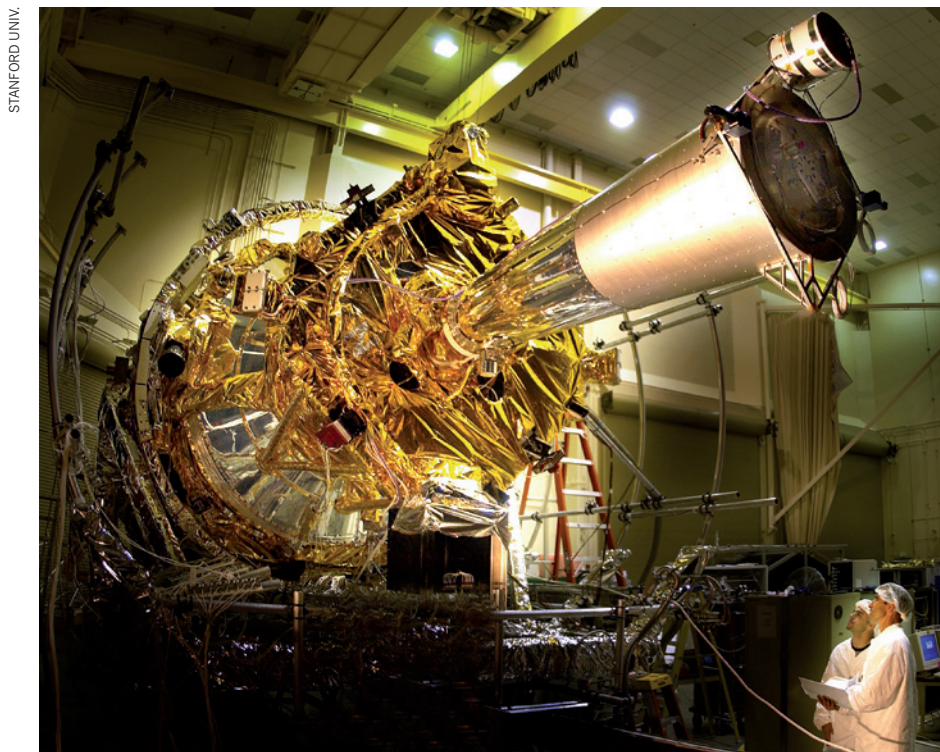
EARTH SCIENCE Has humanity initiated a new geological epoch? **p.133**

ENERGY Report foresees inexorable growth of green energy **p.134**

PHYSICS Much-exaggerated rumours of the Higgs particle **p.136**



COMMUNITY Nature poll: how scientists tend their online reputation **p.138**



By the time Gravity Probe B (above) launched, its goals had largely been met by other projects.

PHYSICS

Troubled probe upholds Einstein

General relativity vindicated, but was the mission worth it?

BY EUGENIE SAMUEL REICH

An epic victory over daunting challenges, or a costly project that should never have flown? After nearly half a century of work and US\$750 million spent, Gravity Probe B, one of NASA's longest-running mission programmes, has finally achieved some scientific closure. But it has yet to quiet its critics.

On 4 May, researchers released the results of a tortuous five-year data analysis that relied on the largesse of a Saudi funding agency to

complete. The verdict, to be published in *Physical Review Letters*: Einstein was right. "I am both glad and relieved that we pulled this off," says physicist Francis Everitt of Stanford University in California, who has led the effort since the beginning.

To some physicists, however, the real impact of Gravity Probe B is to illustrate why future missions should be ranked against competing proposals to improve the scientific return on investment. "I think there are a lot of lessons in this," says Neil Cornish, a physicist at Montana State University in Bozeman, who has

long been sceptical that the mission offered value for money.

First funded in 1963, Gravity Probe B relied on technology that was out of reach for decades. Finally launched in 2004, it carried four gyroscopes — made up of fused quartz balls coated with superconducting niobium that rotated up to 5,000 times per minute. Each ball produced a magnetic field, so that changes in their orientations relative to a guide star, IM Pegasi, could be measured. The point of the measurement was to confirm two predictions of general relativity. One is geodetic precession, in which the curvature of space-time around a massive object, such as Earth, induces a slight wobble in an orbiting gyroscope. Another, much smaller effect is gravitomagnetism, or frame-dragging, in which the spin of a massive object tugs space-time in the direction of its rotation, like a spoon twisted in honey (see 'A twist in space-time'). Gravity Probe B has confirmed the first effect to within 0.3% and the second to within 19%, Everitt says.

But geodetic precession had already been confirmed to nearly this level of accuracy in measurements of laser light bouncing off mirrors on the Moon, and the frame-dragging result is no more accurate than an estimate extracted from measurements of the precession of the orbits of the Laser Geodynamics Satellites (LAGEOS) launched in 1976 and 1992. Furthermore, Gravity Probe B's stated goal was to measure the effects to a precision of 0.01% and 1%, respectively.

Both of Gravity Probe B's results had to be carefully teased out of the data because of unexpected electrostatic effects discovered during the probe's 17-month mission that particularly swamped the frame-dragging signal. "It's a heroic rescue," says Clifford Will of Washington University in St Louis, Missouri, the chair of Gravity Probe B's advisory board.

Others see the complexity of the calculations as a reason to doubt the probe's frame-dragging measurement. "It may be that people repeating this analysis with another working hypothesis on the nature of the systematic errors would get another result," says Ignazio Ciufolini of the University of Salento in Lecce, Italy, who published the results from LAGEOS (I. Ciufolini and E. C. Pavlis *Nature* **431**, 958–960; 2004).

Conscious of previous confirmations of general relativity's effects, NASA convened an external panel in 1995 to assess whether ►

► Gravity Probe B was still worthwhile. The panel gave the probe a green light, but it did not rank it against other possible missions. That was a mistake, says Cornish, because most well-conceived missions look useful when considered in isolation. “It is a waste of time to do reviews of a single mission,” he says. Everitt responds that Gravity Probe B was competitively reviewed numerous times during its development, but he acknowledges that, as a fundamental physics mission, it did not go through the same review process used

to select astronomy missions.

In 2008, when researchers were analysing data from the flight, Cornish served on a review panel that ranked Gravity Probe B as the lowest of eight then-active missions in terms of science per dollar. In response, NASA took the remarkable step of cutting off funding for the probe before its results were in. But Stanford University, the prime contractor on the mission, succeeded in keeping the data analysis alive. Everitt and his team raised an additional \$3.73 million from private sources,

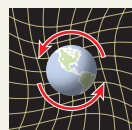
most of it from the Saudi science-funding agency King Abdulaziz City for Science and Technology (KACST), arranged through Turki Al Saud, a Saudi Arabian prince with a PhD from Stanford who is vice-president for research at the KACST.

Gravitation expert Bernard Schutz of the Max Planck Institute for Gravitational Physics in Potsdam, Germany, says he thinks that the mission was worthwhile because general relativity should be checked in a variety of ways. “The fact that they’re getting the same thing is what we want in physics. I think it’s fantastic,” he says. But he laments the delays to launching a dedicated satellite that would amount to a more precise version of LAGEOS, as proposed by Ciufolini in the 1980s. “This could have been done a decade ago,” Schutz says.

The Laser Relativity Satellite (LARES), on which Ciufolini is principal investigator, will finally be launched this year by the Italian Space Agency. The team intends to measure frame-dragging with 1% precision by monitoring the precession of LARES’s orbit. Its cost of about €4 million (US\$5.7 million), not including launch, pales by comparison with that of Gravity Probe B. ■

A TWIST IN SPACE-TIME

Gravity Probe B confirmed key predictions of general relativity by measuring two tiny shifts in the spin axis of its gyroscopes.

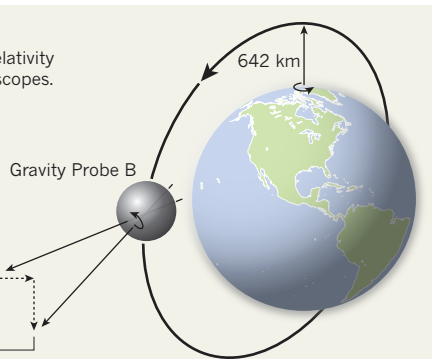


Frame-dragging effect
(0.000011 degrees per year)
Spinning object twists space-time.



Geodetic effect
(0.0018 degrees per year)
Massive object curves space-time.

Guide star
IM Pegasi
(HR 8703)



MICROBIOLOGY

Salmonella hits US teaching labs

Wave of infections triggers investigation into biosafety practices.

BY ERIKA CHECK HAYDEN

A spate of lab-associated *Salmonella* infections has swept across the United States during the past year, prompting public-health officials to examine how closely labs are following infection-prevention protocols.

“The fact that cases seem to be happening all over the country has raised the question of whether there are issues with laboratory safety and appropriate training techniques,” says Mack Sewell, state epidemiologist at the New Mexico Department of Health in Santa Fe.

Between August 2010 and March this year, 73 infections due to *Salmonella typhimurium*, a relatively common strain of the bacterium, caused illness in people across 35 states and one death, the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, reported on 28 April. The outbreak now seems to have ended, with the number of reported new infections dropping to the usual baseline of 0–4 per week, the CDC says.

The strain of *Salmonella* involved in the outbreak often causes food-borne illnesses and has been linked to past epidemics. But what sets this recent outbreak apart is that many of

the illnesses have been traced back to clinical or teaching laboratories, according to the CDC report.

The agency conducted an in-depth investigation of 32 people made ill during the current outbreak, and found that 60% of them had had some connection with a microbiology laboratory in the week before their illness, compared with 2% of 64 people with other reported illnesses. The New Mexico Department of Health found that the outbreak strain was genetically identical to a commercially available strain used in some of the labs linked to ill workers or students. Furthermore, some victims of the illness reported working with *Salmonella* at the bench.

Laboratory-acquired infections are relatively common. One voluntary survey of 88 labs between 2002 and 2004 found that 33% had experienced at least one lab-acquired infection during that time, including six *Salmonella* infections (E. J. Baron and J. M. Miller *Diagn. Microbiol. Infect. Dis.* **60**, 241–246; 2008). In 2008, the CDC convened a panel to look into the incidence of laboratory-acquired *Salmonella* infections; it has not yet released a report.

But it is unusual for lab-acquired infections to crop up across the country, and the CDC

is investigating how this may have happened, in part by surveying the biosafety practices of members of the American Society for Microbiology, based in Washington DC, and the Association of Public Health Laboratories in Silver Spring, Maryland.

The CDC is also concerned because some of those made ill in the outbreak, such as young children, had never visited a lab but lived with someone who worked in a lab and did not get sick. This suggests that the lab worker carried the pathogenic bacteria home on bags, clothes or other objects, says the CDC’s Casey Barton Behravesh. And it raises the question of whether use of this pathogenic *Salmonella* strain in teaching labs is necessary.

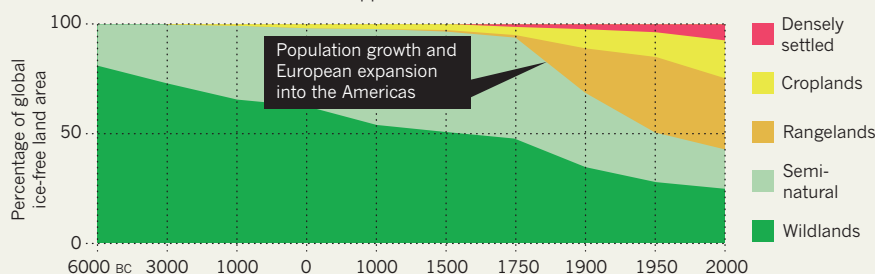
“We are wondering whether there is a non-pathogenic or attenuated strain of *Salmonella* that could be used in place of this one,” Barton Behravesh says.

For now, public-health officials are advising all lab workers to be more careful about observing proper biosafety procedures. Recalling a memorable lesson imparted by his own professor, Sewell suggests one way to cement the importance of those precautions: “He told us that anyone who catches something they’re working with gets an automatic ‘F.’” ■

SOURCE: REF. 1

TRANSFORMATION OF THE BIOSPHERE

The effects of human intervention are now apparent on more than half of Earth's ice-free land mass.



EARTH SCIENCE

Human influence comes of age

Geologists debate epoch to mark effects of Homo sapiens.

BY NICOLA JONES

Humanity's profound impact on this planet is hard to deny, but is it big enough to merit its own geological epoch? This is the question facing geoscientists gathered in London this week to debate the validity and definition of the 'Anthropocene', a proposed new epoch characterized by human effects on the geological record.

"We are in the process of formalizing it," says Michael Ellis, head of the climate-change programme of the British Geological Survey in Nottingham, who coordinated the 11 May meeting. He and others hope that adopting the term will shift the thinking of policy-makers. "It should remind them of the global and significant impact that humans have," says Ellis.

But not everyone is behind the idea. "Some think it premature, perhaps hubristic, perhaps nonsensical," says Jan Zalasiewicz, a stratigrapher at the University of Leicester, UK, and a co-convenor of the meeting. Zalasiewicz, who declares himself "officially very firmly sitting on the fence", also chairs a working group investigating the proposal for the International Commission on Stratigraphy (ICS) — the body

that oversees designations of geological time.

The term Anthropocene was first coined in 2000 by Nobel laureate Paul Crutzen, now at the Max Planck Institute for Chemistry in Mainz, Germany, and his colleagues. It then began appearing in peer-reviewed papers as if it were a technical term rather than scientific slang.

The "evidence for the prosecution", as Zalasiewicz puts it, is compelling. Through food production and urbanization, humans have altered more than half of the planet's ice-free land mass¹ (see 'Transformation of the biosphere'), and are moving as much as an order of magnitude more rock and soil around than are natural processes². Rising carbon dioxide levels in the atmosphere are expected to make the ocean 0.3–0.4 pH points more acidic by the end of this century. That will dissolve light-coloured carbonate shells and sea-floor rocks for about 1,000 years, leaving a dark band in the sea-floor sediment that will be obvious to future geologists. A similar dark stripe identifies the Palaeocene–Eocene Thermal Maximum about 55 million years ago, when global temperatures rose by some 6°C in 20,000 years. A similar temperature jump could happen by 2100, according to some high-emissions scenarios³.

The fossil record will show upheavals too. Some 20% of species living in large areas are now invasive, says Zalasiewicz. "Globally that's a completely novel change." And a review published in *Nature* in March⁴ concluded that the disappearance of the species now listed as 'critically endangered' would qualify as a mass extinction on a level seen only five times in the past 540 million years — and all of those mark transitions between geological time periods.

Some at the ICS are wary of formalizing a new epoch. "My main concern is that those who promote it have not given it the careful scientific consideration and evaluation it needs," says Stan Finney, chair of the ICS and a geologist at California State University in Long Beach. He eschews the notion of focusing on the term simply to "generate publicity".

Others point out that an epoch typically lasts tens of millions of years. Our current epoch, the Holocene, began only 11,700 years ago. Declaring the start of a new epoch would compress the geological timeline to what some say is a ridiculous extent. Advocates of the Anthropocene, however, say that it is natural to divide recent history into smaller, more detailed chunks. A less controversial alternative would be to declare the Anthropocene a new 'age': a subdivision of an epoch.

If scientists can agree in principle that a new time division is justified, they will have to settle on a geological marker for its start. Some suggest the pollen of cultivated plants, arguing that mankind's fingerprint can be seen 5,000–10,000 years ago with the beginnings of agriculture. Others support the rise in the levels of greenhouse gases and air pollution in the latter part of the eighteenth century, as industrialization began. A third group would start with the flicker of radioactive isotopes in 1945, marking the invention of nuclear weapons.

Should the working group decide that the Anthropocene epoch has merit, it will go to an ICS vote. But the whole process will take time — defining other geological periods has sometimes taken decades. In the meantime, Zalasiewicz says, "the formalization is the excuse to try to do some very interesting science", comparing Earth's current changes to those of the past. ■

1. Ellis, E. C. *Phil. Trans. R. Soc. A* **369**, 1010–1035 (2011).
2. Zalasiewicz, J., Williams, M., Haywood, A. & Ellis, M. *Phil. Trans. R. Soc. A* **369**, 1036–1055 (2011).
3. IPCC *Climate Change 2007: The Physical Science Basis* (IPCC, 2007).
4. Barnosky, A. D. *et al. Nature* **471**, 51–57 (2011).



TOP NEWS



Nuclear power plants cleared of leukaemia link go.nature.com/u2ehtq

OTHER NEWS

- Fracturing rock to extract natural gas raises methane in drinking water go.nature.com/zbas2o
- How to get funded when success rates fall go.nature.com/qytird
- Australia's new chief scientist reveals his plans go.nature.com/qgqut1

PODCAST



In the backyard: new type of fungus found in university pond go.nature.com/xtrlak

VENTURE MEDIA GROUP/AURORA PHOTOS/CORBIS

ENERGY

How green is my future?

UN panel foresees big growth in renewable energy, but policies will dictate just how big.

BY JEFF TOLLEFSON

For centuries, humans have powered a growing world by extracting Earth's carbon-rich rocks, peat and liquids and burning them in ever greater amounts, but that trend is beginning to change. A report from the Intergovernmental Panel on Climate Change (IPCC), released on 9 May at a briefing in Abu Dhabi, suggests that an inevitable — if slow — shift towards specialized energy crops, sunlight, wind and other sources of renewable power will mark the next four decades.

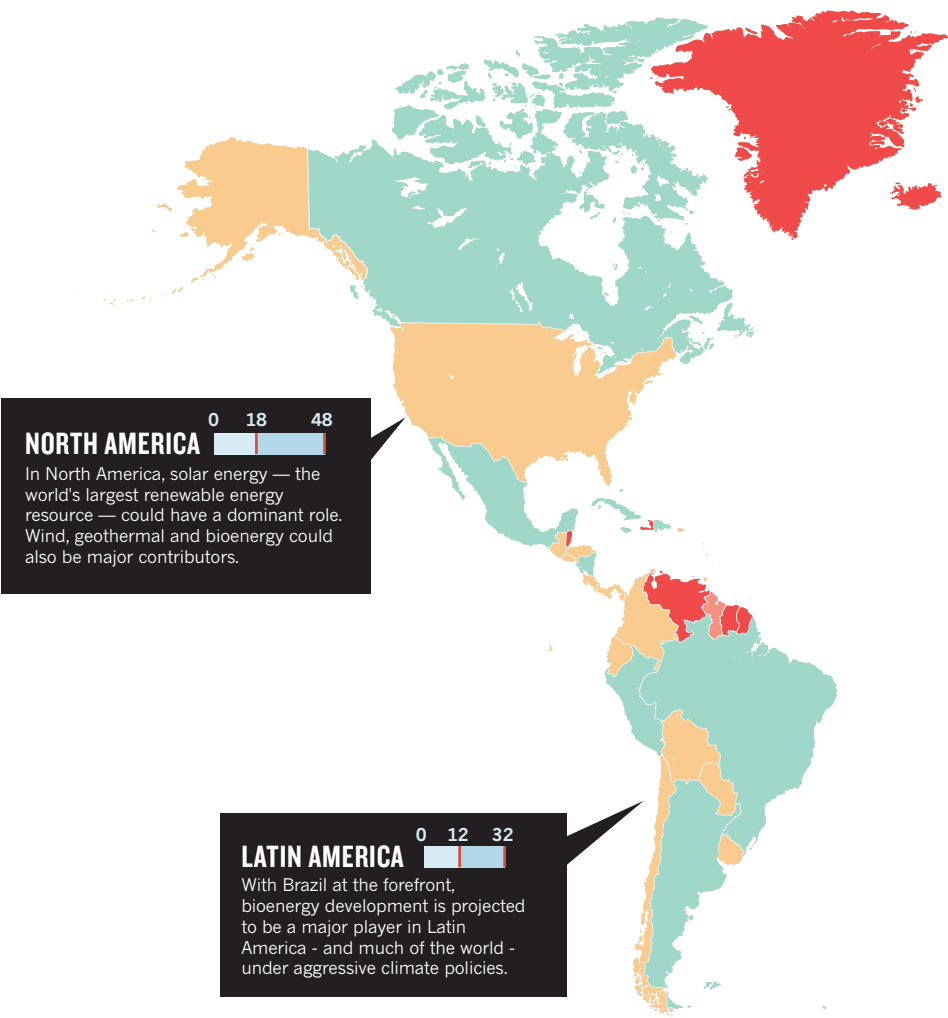
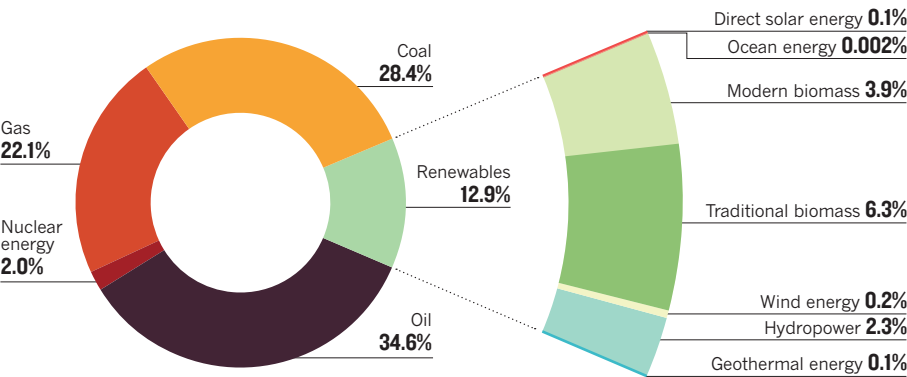
In addition to surveying published work on renewable energy potential, production, economics and policy, the IPCC conducted detailed socioeconomic modelling across 164 scenarios, both with and without policies intended to reduce greenhouse-gas emissions. The extent and type of development varies significantly depending on factors such as price and technological progress, and on government policies intended to reduce greenhouse-gas emissions. “Renewables will have a great future even without climate policies, but that does not necessarily lead to an emissions reduction,” says Ottmar Edenhofer, chairman of the IPCC's working group on mitigation.

Renewable energy sources, excluding the burning of traditional biomass such as wood, make up roughly 7% of global energy production (see ‘A small slice’), and the working group's *Summary for Policymakers* estimates that by 2050, production could rise to between three and more than ten times its current level. No renewable technology is projected to dominate, but bioenergy, solar energy and wind will be important in the energy mix (see ‘The big three’). This shift towards renewable energy could reduce cumulative greenhouse-gas emissions by 220–560 gigatonnes of carbon dioxide, from a baseline estimate of 1,530 gigatonnes.

In all scenarios, fossil fuel will continue to have an important role for decades to come. The report says that falling prices for renewable energy will drive a gradual deployment of clean energy, but overhauling the global energy system represents a monumental task (see ‘Planet renewable’). The IPCC says there are no technical barriers preventing large-scale deployment of renewable-energy systems over the next few decades. And even in 2050, none of the four main scenarios shows humanity tapping more than 2.5% of the accessible supply of renewable energy. ■

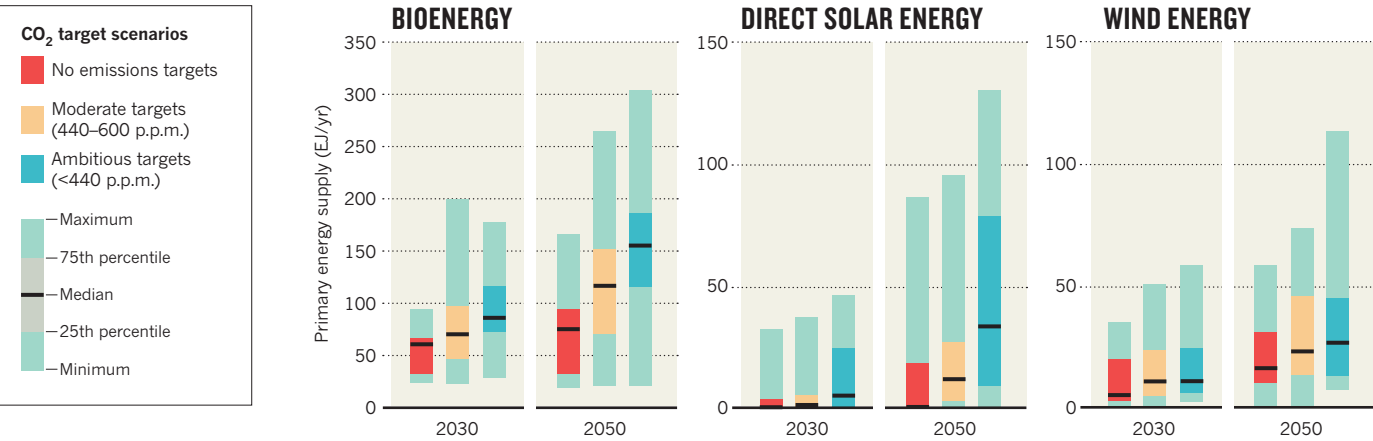
A SMALL SLICE

The world consumed around 492 exajoules (10¹⁸ joules) of energy in 2008, and renewable energy made up 12.9% of the total, and less than 7% if the burning of traditional biomass is excluded.



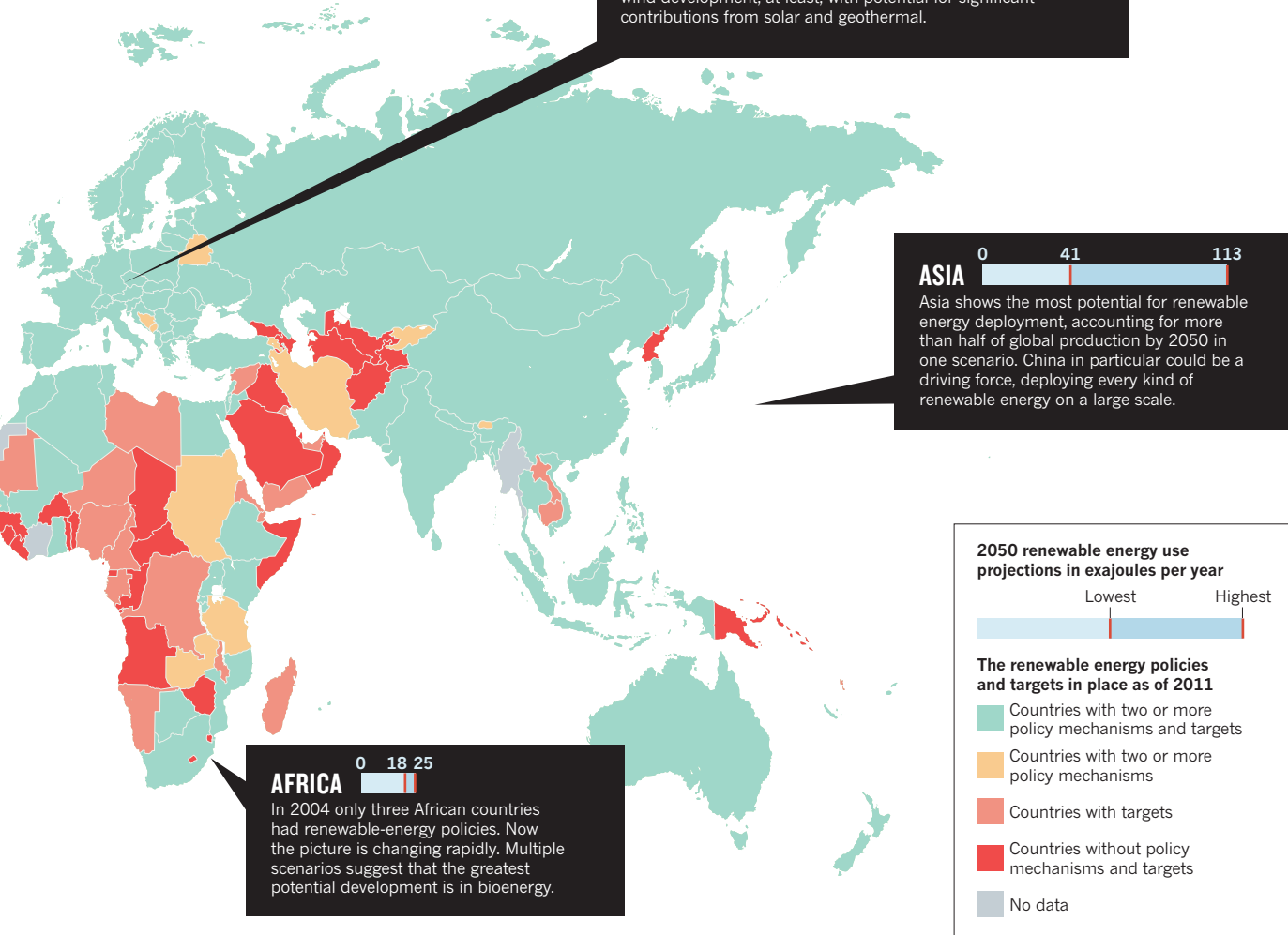
THE BIG THREE

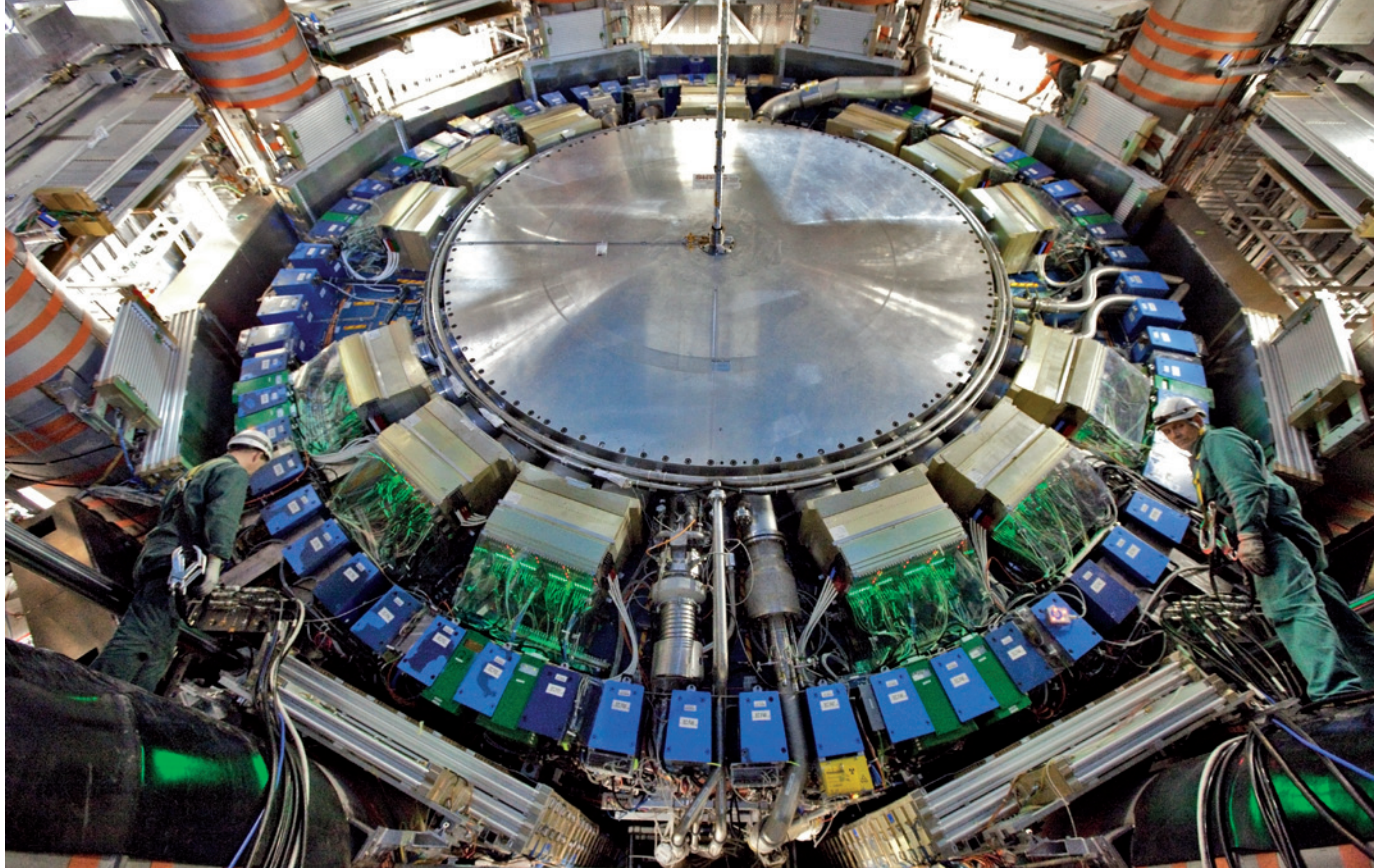
Under the emissions scenarios studied by the Intergovernmental Panel on Climate Change, future renewable energy development varies with the level of ambition in halting the rise of atmospheric carbon dioxide concentrations. These rose to more than 390 parts per million (p.p.m.) in 2010. Biomass, solar energy and wind power look to remain the key renewable energy types.



PLANET RENEWABLE

Four scenarios representing a range of policies and conditions show that by 2050 renewables could grow from 2008 levels of 32 exajoules per year to about 100–300 EJ/yr globally. Renewable energy could provide 15–77% of the global power supply. The lowest and highest projections in the four scenarios are estimated for selected regions.





C. MARCELLONI/CERN

The ATLAS experiment at the Large Hadron Collider: jumping at shadows?

PARTICLE PHYSICS

The collider that cried 'Higgs'

Data leaks from particle hunters raise questions about scientific trust.

BY GEOFF BRUMFIEL

In the era of WikiLeaks and Twitter, can anyone keep a secret? Governments have learned that, all too often, the answer is no. Now, as teams of particle physicists close in on one of their biggest targets in decades, they too are struggling to keep confidential data under wraps.

In late April, leaked results from the Large Hadron Collider (LHC), the world's largest particle accelerator, seemed to show a preliminary signal of the Higgs boson. The particle is the LHC's highest-profile quarry, and would provide evidence of a theoretical mechanism that gives other particles their mass. A fresh analysis published this week has debunked the claim, but researchers are bracing themselves for a string of other false alarms to appear on blogs over the coming months.

Future leaks are "inevitable", says James Gillies, a spokesman for CERN, the European particle-physics laboratory near Geneva, Switzerland, where the LHC is housed. Proof of the Higgs' existence will not arrive as a bolt from the blue — instead, it will emerge slowly from weeks or months of data analysis, allowing ample time for each tantalizing step to be documented on blogs.

Rumours about particle discoveries are as old as the field itself, but the ease of communicating tentative finds via the Internet means that particle hunters' first clues and false starts are now being discussed on the global stage. This raises broader questions for a discipline that employs thousands of scientists to search for a handful of discoveries worthy of a Nobel prize. With the stakes so high, can preliminary results be kept secret? Should leakers or bloggers be punished for making early findings public? Above all, says Adam Falkowski, a blogger and theorist at the University of Paris-South: "Is this bad, or not?"

The trouble began shortly after 6:45 p.m. Geneva time on 21 April, when someone posted an internal communication from the LHC's ATLAS particle detector to the popular blog 'Not Even Wrong' (go.nature.com/q55vdu). The note purported to show preliminary evidence for γ -rays from a decaying Higgs particle with a mass of about 115 gigaelectronvolts, some 115 times heavier than a proton. The number was familiar to many CERN physicists, who had seen hints of the Higgs at that mass in a previous accelerator, the Large Electron-Positron Collider.

"Higgs? That looks very, very interesting!" wrote Peter Woit, a mathematician at Columbia University in New York City who runs Not Even Wrong. Anonymous physicists working at CERN quickly confirmed in the blog's comments thread that the data were genuine, but said questions remained about the analysis.

News of the document spread first to other blogs, then to science publications, and finally to the mainstream media. But as the news gathered momentum, it became clear to ATLAS researchers that it was a false alarm. By 24 April, Jonathan Butterworth, a physicist at University College London who works on ATLAS, was looking sheepish on British television news. "What's happened here is some people have spent four nights without sleep and got rather overexcited," he explained.

"I was surprised by the level of interest," Butterworth told *Nature* last week. He is relaxed about the leak, seeing it as an opportunity to discuss scientific process with the public. But others in the collaboration are less pleased. "It got me quite angry," says Gordon Watts, an ATLAS physicist at the University of Washington in Seattle. Watts says that the leak forced many in the group to spend their Easter holiday scrutinizing the new result. An official ATLAS analysis refuting the claim appeared on

NATURE.COM
For *Nature's* special
on the LHC see:
www.nature.com/lhc

8 May (go.nature.com/zdzlji), but has not yet been widely reported in the mainstream media.

Finding out who posted the document, and why, may be impossible. On Not Even Wrong, some accused the lead author on the internal communication, physicist Sau Lan Wu of the University of Wisconsin, Madison, of seeking Higgs glory. She flatly denies that she had anything to do with the leak. "It's absolutely not from me or my group," she says. "I have never gone to a blog." Other commentators suggest that the post was actually an attempt to make Wu look bad: ATLAS explicitly prohibits its collaborators from publicly distributing or discussing internal documents or results before formal publication.

Woit says that he cannot trace the Internet address of the anonymous leaker. Even if he could, he says, "I just don't see an argument that I should be enforcing ATLAS's privacy policies".

FLOOD OF LEAKS

Perhaps inevitably, the leak was soon followed by a counter-leak. On 4 May, *New Scientist* magazine reported that ATLAS's main rival, the Compact Muon Solenoid (CMS) detector, had seen no evidence of the earlier leaked signal. Guido Tonelli, the spokesman for the CMS group, says that the report was probably based on slides from a presentation made a few days earlier.

Tonelli says that even if he found out who leaked the CMS data, he would do little more than talk to them about the issue. He acknowledges that future leaks are probably unavoidable. All 3,000 members of the CMS collaboration must have access to the data, both out of fairness and to ensure the accuracy of new findings. "I might be able to reach an efficiency of 99.9% in convincing the collaboration" to obey the rules, Tonelli says. "But three people are still enough to create a leak."

Others are divided over just how damaging the leaks are. "I don't really see damage; I see more inconvenience," says Falkowski. He believes that the press and public were quick to recognize that the leak was not an official finding. Watts nevertheless feels that it may have damaged the reputation of the collaboration and the field. "It's a credibility thing in the end," he says.

The debates, like the leaks, are sure to continue. Many physicists hope to find the Higgs hiding at around 115–130 gigaelectronvolts, where its signal can be confused with those from other particle decays. As the LHC explores that energy range, the uncertain data will mean plenty of false alarms. And as Butterworth tweeted to his followers: "If u get excited by unreviewed results leaked in breach of confidence you'll have a distracting few months." ■

OCEANOGRAPHY

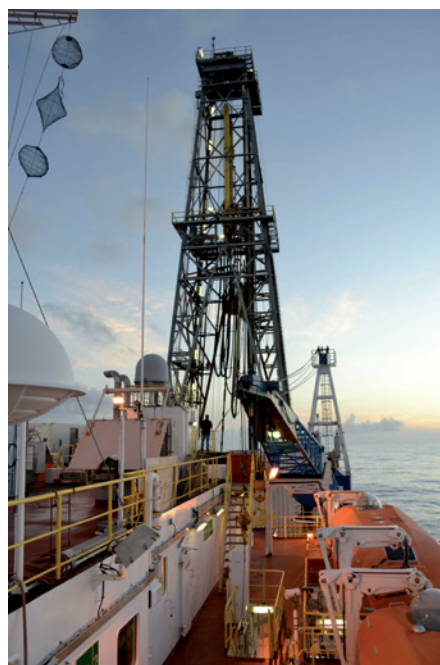
Costs to keep ocean drilling ship in port

Ten-year programme aims for rebirth in tough climate.

BY NICOLA JONES

An international effort to explore the little-known depths below the ocean floor is heading into choppy waters. Last week, the US branch of the 24-nation Integrated Ocean Drilling Program (IODP) received guidance on its budget allocation from the US National Science Foundation, one of the project's two main funders (the other being the Japan Agency for Marine-Earth Science and Technology). No cuts are anticipated to the US share of the project's budget, but a rise in fuel costs of more than 30% since the end of 2010 means that in the 2012 fiscal year the US drilling ship the *JOIDES Resolution* is expected to be at sea for only six months. It needs eight months to complete its four scheduled expeditions, so one project — to drill into the tectonic area off the southern coast of Alaska next summer — has been postponed.

In an open letter to colleagues last week, David Divins, director of ocean drilling programmes at the Consortium for Ocean Leadership in Washington DC, which helps to run the US component of the IODP, wrote, "I am extremely troubled by this news and the implications for the future of scientific ocean drilling."



The *JOIDES Resolution* will spend less time at sea.

News of the vessel's scaled-back season comes just as the IODP sets its scientific goals for its next ten years. The plan is expected to be finalized this month and released in June. It highlights an ongoing emphasis on the studies of geophysics and past climates that has dominated the programme's initial ten-year mandate, which runs until 2013. "Everything we know about climate older than a million years we know from the ocean drilling programme," says Maureen Raymo, chair of the IODP's main science evaluation committee.

The next phase will also up the emphasis on studies of deep biology, and will include more efforts to install instruments in boreholes to conduct real-time experiments, rather than simply pulling up cores. And this month, scientists will be meeting in Tokyo to discuss the possibility of drilling into the fault line of the great earthquake that hit Japan in March. If they can do so within the next two years, they should be able to detect the heat generated by the slip, a clue to energy transfer during quakes.

But logistical challenges have hobbled the IODP in the past. The *JOIDES Resolution* was out of action for nearly three years for refurbishments that took much longer than anticipated. A new Japanese vessel capable of drilling to extreme depths, the *Chikyu*, has done only about 14 months of active science service since mid-2007. It was also damaged by Japan's recent tsunami, causing the cancellation of one expedition.

As a result, and in the face of serious technological challenges, the IODP has been slow to meet its original objectives. Only two expeditions have been launched specifically to study the biosphere deep beneath the sea floor, a key goal. Meanwhile, just one deep core has been retrieved from the Arctic, another major target, and investigators are behind schedule on drilling into a subduction zone — where an oceanic plate slides beneath a continental one — to study earthquake mechanisms.

Ongoing battles over US government spending have some in the drilling community concerned that the programme's cost will make it a tough sell going forwards. Divins estimates that, in total, the four expeditions originally planned for the 2012 financial year would have cost about US\$70 million. "We have to make a compelling case to keep this ship out there," says Raymo. ■

E. GANBAT



BEST FACE FORWARD

A Nature poll reveals how researchers guard, and sometimes burnish, their online image.

BY EUGENIE SAMUEL REICH

In March, a puzzling press release began circulating around the Internet. Titled “DR. Anil Potti Likes Spending Quality Time With His Wife And Three Daughters”, it listed a series of qualifications, honours and prizes won by Potti, a cancer geneticist formerly at Duke University in Durham, North Carolina. He resigned in 2010 after it was revealed that he had falsely claimed to be a Rhodes scholar and Duke began to investigate errors in his work.

Most scientists embroiled in scandal shrink from view, but Potti’s online presence began booming in unexpected ways. After he resigned, he, or someone using his name, created more than half a dozen websites about him and his research, including: www.pottianil.com; www.anilpotti.com; www.anilpotti.net; and www.dranilpotti.com. Twitter and Facebook accounts appeared in his name in January, followed by a stream of press releases notable for their breathless banality: aside from enjoying time with family, Potti believes that most lung cancers are caused by smoking; he is an advocate for personalized cancer therapy;

and he donates his time and money to the local school system and church. It is difficult to identify the source of the releases, websites and social-media accounts. Potti could not be reached for comment, and his lawyer, Jim Maxwell of Maxwell, Freeman & Bowman in Durham, declined to comment on them, explaining that there is a confidential research-misconduct investigation ongoing at Duke. “Until that is concluded and Dr Potti is cleared of any wrongdoing, he is not in a position to be making public comments,” Maxwell says.

The only clue to their origin lies in the registration information for www.anilpotti.com and www.anilpotticv.com, which list an e-mail address from Online Reputation Manager as an administrative contact.

Online Reputation Manager, headquartered near Rochester, New York, is a company that uses search-engine optimization strategies to repair the online image of clients who have been besieged with unfavourable press. These include flooding the Internet with positive messages to drown out the negative. A company representative confirmed ownership of

the e-mail address, but could not say whether Potti is a client.

Potti’s reputation may have needed serious work, but a *Nature* poll reveals that a significant number of scientists are concerned with maintaining their online image. The poll was e-mailed to 30,000 working scientists and was promoted on Facebook and Twitter. Of 840 respondents, 77% say that their personal online reputation is important to them and 88% say that the online reputation of their work is important (see ‘A name online’). Thirteen per cent say that they have used search-optimization strategies to improve the visibility of their research, and as many as 10% say they have considered using external services to manage their online reputations.

Several researchers have set up biographies on the online site Wikipedia — the online encyclopaedia that practically anyone can edit — or edited entries to include references to their own papers. And many simply use social-networking sites or blog regularly about science, which can help to shape a digital persona. The poll and subsequent interviews suggest a growing recognition in the scientific community: maintaining a prominent online presence can help researchers to network with colleagues, share resources, raise money and communicate their work. “It is incredibly valuable,” says Gia Milinovich, a web producer based in London who has studied scientists’ use of Twitter.

At a minimum, says Alex Bateman, a bio-informatician at the Wellcome Trust Sanger Institute near Cambridge, UK, scientists should ensure that they have an online profile that includes contact details. For his part, he routinely checks that his publications come up together in a list when his name is searched in databases such as Web of Science and Scopus. If he finds an error, he contacts the database company to complain. “They’re very quick to respond,” he says.

Others are looking at the face they present to the wider world, through sites such as Wikipedia. For many people looking for information on a scientific topic, Wikipedia is a first port of call — and our poll shows that scientists use it regularly. As many as 72% admit to checking Wikipedia at least once a week and about one-fifth do so for references to themselves or their group’s work. Nine per cent of our sample say that they have inserted references to their or their group’s work in the past 12 months, and nearly 3% have edited their biographies, something that is frowned on by Wikipedia editors, according to Bateman. “You shouldn’t be editing articles you are too close to because you have a conflict of interest,” he says.

That said, roughly one-tenth of the respondents to our poll say that their work has been misrepresented on the web, and some scientists in this situation feel the need to set the record straight. Walt de Heer, a physicist at the Georgia Institute of Technology in Atlanta, works

ILLUSTRATION BY JOE MAGEE

on graphene — two-dimensional carbon sheets that may have applications in electronics. In 2009, de Heer caught wind of rumours that his research had been inspired by work done at the University of Manchester, UK, by Andre Geim and Konstantin Novoselov (the pair won the Nobel prize for this research in 2010). De Heer saw that the Wikipedia article on graphene emphasized the Manchester work and suspected that it was fuelling the rumours. So de Heer created his own biography on Wikipedia. And although it was nominated for deletion by at least one Wikipedia editor, enough users of the site have agreed that it should remain.

Darren Logan, a geneticist at the Wellcome Trust Sanger Institute, is an administrator on Wikipedia — a position that gives him additional editing powers. He agrees that editing Wikipedia can be a very influential way of getting a point across, even within the scientific community. One article he has written, on major urinary proteins, included references to his scientific work and introduced terminology that others later used to describe his work. “The purpose of writing wasn’t to promote my own work but a consequence was that a lot more people read my research articles. It’s influencing them,” he speculates.

MANIPULATING VISIBILITY

A handful of researchers are using more sophisticated tools to increase the visibility of a website. Software engineer Brian Turner has been trying to promote software developed at a lab at the Hospital for Sick Children in Toronto, Canada, where he works. He uses Google’s webmaster and analytics tools to figure out how Google ‘sees’ the lab’s website and how much traffic the website gets through

Google. The analysis prompted him to change the titles of several pages from obscure identifiers to ones that include the names of proteins that people might search for. “That made a big difference to our search rankings,” he says.

Social-networking tools can also boost a person’s visibility on the web. Among the 549 people who responded to the e-mail invitation to take part in our survey, 59% had used Facebook and 23% had used Twitter. About 17% of them had written at least one blog. Although blogging is usually deemed extra-curricular, some say it has definite career benefits. Paul

Nearly 3% of respondents had edited their own Wikipedia biographies.

‘P. Z.’ Myers, a biologist at the University of Minnesota Morris, runs the popular blog Pharyngula, which he says gets about a million visits a month. He says that he has never mentioned the blog on his CV or applications for tenure, but his tenure referees raised the blog as an example of something positive he was doing.

Unsurprisingly, younger researchers tend to be more preoccupied with online reputations than older ones. Although more than half of researchers under 35 say that they strongly agreed that the online reputation of their work was important, that number dropped to 42% for those aged 35–54 and to 32% for those aged 55 and over. Peter Ruben, a biophysicist at Simon Fraser University in Burnaby, Canada,

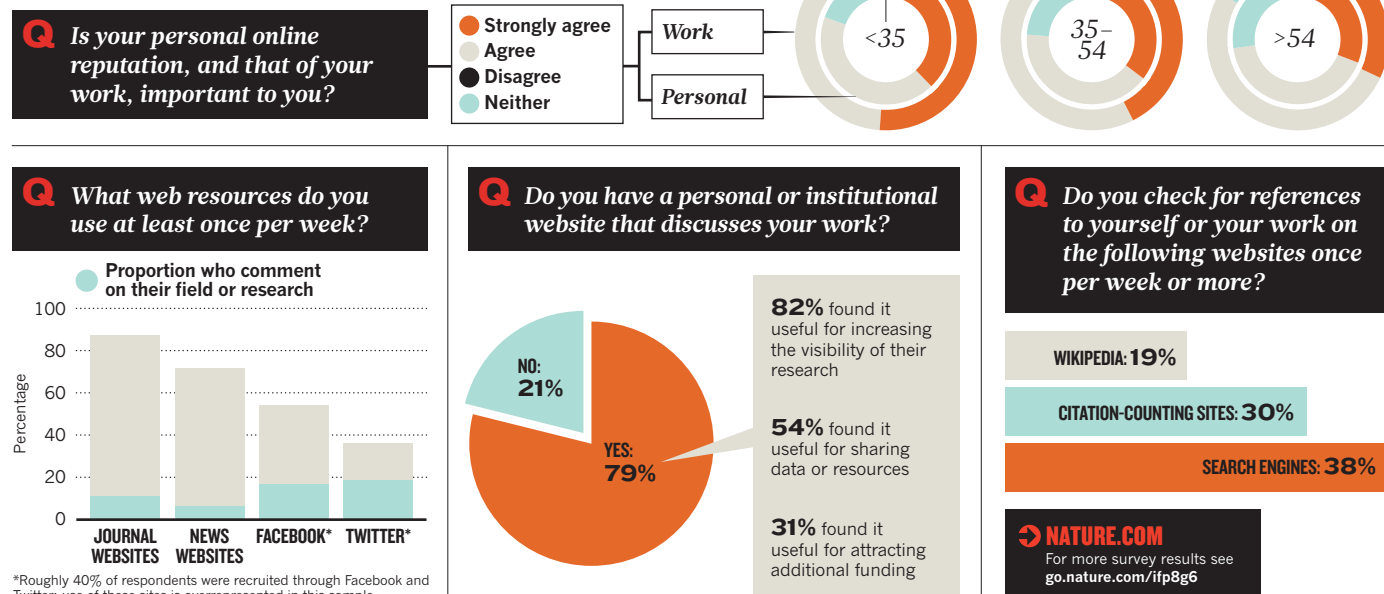
falls in that latter demographic and says he doesn’t care about his online reputation. In 2005, Ruben published a paper (S. L. Geffeney *et al.* *Nature* **434**, 759–763; 2005) that reported on evolution of resistance to a toxin in garter snakes. Although his work was misrepresented on creationist websites, Ruben didn’t try to set the record straight, and doesn’t really think that it has tarnished his reputation.

The positive messages being posted on Potti’s behalf have had some effect. In a Google search for ‘Anil Potti’ on 9 May, five of the top ten links were to positive material placed in the past several months. But a detractor has surfaced, setting up a satirical Twitter account, @anil_potti, which posts links to articles about the ongoing investigation at Duke. An article in the independent student newspaper, *The Chronicle*, also questioned the ethics of the online management activity, pointing out that it discusses Potti’s research without saying that it has been questioned. Ronald Smith, manager of business development at Online Reputation Manager, stresses that the kind of work it does is ethical, legal and accepted in the search-optimization industry. But Bateman says that whoever is doing the work on Potti’s reputation has their work cut out for them. “In the Internet world it is impossible to remove the evidence. In this case Anil Potti is in danger of bringing more attention to his alleged scientific misconduct.” Potti’s is an extreme case, but it holds a lesson for anyone who sets out to court attention online: the web can be unpredictable. ■ [SEE EDITORIAL P.124](#)

Eugenie Samuel Reich is a reporter for *Nature* in Cambridge, Massachusetts. Survey work was aided by Laura Harper.

A NAME ONLINE

A poll of 840 researchers recruited by e-mail and through social-networking sites reveals how scientists manage their online reputations.





Vesuvius last awoke with a small blast in 1944. A large eruption could unleash incendiary avalanches and ash that would threaten millions of people.

EUROPE'S TICKING TIME BOMB

Vesuvius is one of the most dangerous volcanoes in the world — but scientists and the civil authorities can't agree on how to prepare for a future eruption.

BY KATHERINE BARNES

It starts with a blast so strong that a column of ash and stone rockets 40 kilometres up into the stratosphere. The debris then drops to Earth, pelting the surface with boiling hot fragments of pumice and covering the ground with a thick layer of ash. Roofs crumble and vehicles grind to a halt. Yet the worst is still to come. Soon, avalanches of molten ash, pumice and gas roar down the slopes of the volcano, pulverizing buildings and burying everything in their path. Almost overnight, a packed metropolis becomes a volcanic wasteland.

This is Naples, Italy, in the throes of a cataclysmic eruption of Vesuvius — the volcano that destroyed the city of Pompeii in AD 79. The scenario may sound far-fetched, but in the wake of Japan's recent earthquake and tsunami, many areas are reassessing the risks from their own 'black swans', a term used to describe unlikely but potentially devastating disasters. And Naples stands out as particularly vulnerable, with a population of 3 million living in the shadow of Vesuvius.

The volcano has been eerily dormant since a

small eruption in 1944, but recent studies suggest that Vesuvius could be more dangerous than previously assumed, which has prompted a vigorous debate about the risk and scale of future disasters. Local authorities face the difficult task of deciding how to protect a large population in the event of earthquakes and other signs heralding the volcano's reawakening. "There would be no modern precedent for an evacuation of this magnitude," says Giuseppe Mastrolorenzo at the Vesuvius Volcano Observatory in Naples. "This is why Vesuvius is the most dangerous volcano in the world."

RUMBLINGS OF DISSENT

The slumbering giant won't stay quiet forever. Seismic imaging studies have detected an unusual layer about 8–10 kilometres deep under the mountain's surface. Mastrolorenzo and his colleague Lucia Pappalardo

interpret this layer as an active magma reservoir¹, which could produce large-scale 'plinian'-style explosions — named after Pliny the Younger, who described the AD 79 eruption.

The first rumblings of activity at Vesuvius could come weeks to years before an eruption, but there might be little, if any, warning of the eruption itself. Pappalardo and Mastrolorenzo analysed the geochemistry of rock fragments from past eruptions, and found evidence that magma ascended rapidly — in just a few hours — from its deep chamber to the surface.

For many years the largest known eruption of Vesuvius was that of AD 79. But in 2006, Mastrolorenzo and Michael Sheridan at the University at Buffalo in New York described geological evidence for a much larger blast, about 3,800 years ago in the Bronze Age². Fiery avalanches of ash and debris called pyroclastic flows travelled 20 kilometres and covered the whole of the area of present-day Naples. "The deposits right in the centre of Naples are 4 metres thick," says Sheridan. "Even a few inches would be enough to kill everyone."

► NATURE.COM
For more on volcanic eruptions see:
go.nature.com/Gikjak

BETTMANN/CORBIS

Given these concerns, the Vesuvius observatory team has urged the Neapolitan authorities to base their emergency plan on a worst-case 'maximum possible' eruption similar to the Bronze Age blast. "A crisis could start today," says Mastrolorenzo. "The trouble is that nobody would be able to tell how long it would last, what type of eruption it would be, or how the event would evolve." The researchers recommend the complete evacuation of an area 20 kilometres around Vesuvius if earthquakes and other signs of unrest hint that it is coming back to life.

Not all scientists share this doom-laden outlook. Some groups have even proposed that Vesuvius is becoming less explosive. Bruno Scaillet and his colleagues at the University of Orleans in France argue that the eruptive style of Vesuvius has changed as the magma chambers feeding the eruptions have migrated upwards, with the small 1944 eruption coming from a relatively shallow level 3 kilometres below the surface³. Evidence suggests the magma stored there is less viscous, so it is less prone to causing large explosions. If the past trend holds, says Scaillet, the next eruption could be similar to the most recent ones.

Scaillet adds that the seismically unusual layer 10 kilometres below the surface could be magma, but it could also be some other fluid such as water or brine. "These various issues are far from being settled," he says.

EMERGENCY PLANNING

With the size of any future eruption in doubt, and a public more concerned about day-to-day problems such as traffic and crime, mitigating the hazard of Vesuvius is an enormous task shared by researchers and the civil authorities.

Scientists keep constant tabs on Vesuvius through a network of sensors that monitor for earthquakes, ground deformation and changes in the chemistry of escaping gases.

And Italy's Department of Civil Protection (DPC) maintains a National Emergency Plan for Vesuvius. The plan, first developed in 1995, is based on a scenario for an intermediate-sized eruption, similar to one that occurred

in 1631. That sub-plinian blast killed 6,000 people but affected an area much smaller than the earlier plinian eruptions.

The plan divides the area around the volcano into three regions according to the type of hazard expected. The red zone, closest to Vesuvius, is deemed most at risk from pyroclastic flows, so the plan calls for the evacuation of all 600,000 residents in this area before an eruption starts (see 'In the line of fire'). The main danger in the yellow zone comes from falling ash and small rocks. Officials would wait until the eruption starts, and the wind direction is known, before ordering an evacuation of regions in yellow zones downwind of the volcano. The blue zone is at risk from floods and mud flows triggered by the eruption, and would be evacuated according to the same plan. The city of Naples was excluded from any of the hazard zones because the prevailing wind typically blows ash to the east, away from the city.

In 2003, the DPC announced that it would constantly update the emergency plan to take account of new scientific information. The red zone is being expanded to include the eastern districts of Naples and officials reduced the evacuation time from two weeks to 72 hours, recognizing that there may be less of a warning before the eruption.

Nevertheless, some researchers argue that the plan has ignored important scientific evidence. Last year, Mastrolorenzo and Pappalardo⁴ and Giuseppe Rolandi⁵ of the University of Naples found that even with an intermediate-sized eruption, pyroclastic flows would threaten several municipalities not currently included in the red zone. Mastrolorenzo says that officials should also not wait to evacuate the yellow zone, because fine ash would rapidly fill the air and plunge the area into total darkness. "You have to get people out before it starts," he says. And the wind does sometimes blow towards Naples, so the authorities cannot rule out heavy ashfall in the city, say both Mastrolorenzo and Rolandi.

Putting all the evidence together, they and other researchers insist that the emergency plan should correspond to the 'worst-case

scenario', which means including metropolitan Naples and its 3 million inhabitants.

That makes sense for planning, says Jonathan Fink, a volcanologist at Portland State University in Oregon. Once the volcano shows signs of unrest, authorities and scientists can re-evaluate. "If there is an error on the high side, there is less lost than would be the case in the opposite situation," he says.

In a written response to *Nature*, the DPC advocates evaluating the eruption risk "on the basis of the present state of the volcano and not simply assuming the largest eruption event that ever occurred in the volcanic history". Some scientists agree. "You can't spend [everything]

on the absolute worst case. You need to reduce the risk in a rational way," says Warner Marzocchi at the National Institute for Geophysics and Volcanology (INGV) in Rome. A complete evacuation of Naples' 3 million residents, he says, "would be impossible to manage".

Marzocchi and other researchers are developing modelling tools — based on the probabilities of different scenarios — that could help civil authorities evaluate the evidence during a crisis and choose a course of action. Peter Baxter, an expert in emergency planning at the University of Cambridge, UK, specializes in the impacts of volcanic eruptions and used this type of method successfully during the 1997 eruption in Montserrat in the Caribbean to predict which regions would be affected. A complete evacuation of the island was avoided.

For Vesuvius, Baxter and his colleagues have used geological data and models of eruptive processes to develop an 'event-tree' to display the full range of possible eruptions⁶. If sensors on the volcano pick up signs of magmatic unrest, the analysis suggests a 70% probability of an explosive eruption but only a 4% chance of a catastrophic plinian one. The most likely event is a violent but smaller blast, like the one in 1944, with lava flows and moderate ash emissions.

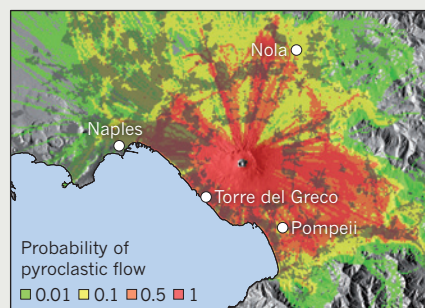
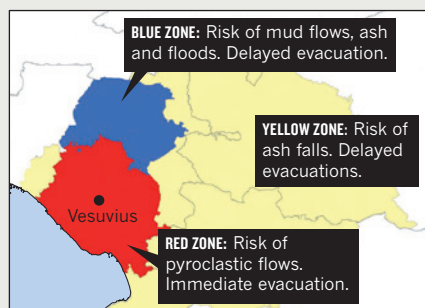
For now, this kind of probabilistic approach seems the only way forward for volcanologists and disaster planners, as there is no recipe for accurate eruption forecasting on the horizon. "It's an extremely complex problem to solve," says Augusto Neri of the INGV's laboratories in Pisa. "We simply do not know how the volcano works." ■

Katherine Barnes is a freelance writer in London.

1. Pappalardo, L. & Mastrolorenzo, G. *Earth Planet. Sci. Lett.* **296**, 133–143 (2010).
2. Mastrolorenzo, G. *et al. Proc. Natl Acad. Sci. USA* **103**, 4366–4370 (2006).
3. Scaillet, B., Pichavant, M. & Cioni, R. *Nature* **455**, 216–219 (2008).
4. Mastrolorenzo, G. & Pappalardo, L. *J. Geophys. Res.* **115**, B12212 (2010).
5. Rolandi, G. *J. Volcanol. Geotherm. Res.* **189**, 347–362 (2010).
6. Baxter, P. J. *et al. J. Volcanol. Geotherm. Res.* **178**, 454–473 (2008).

IN THE LINE OF FIRE

Plans call for initial evacuation of only the zone closest to Vesuvius (left map, red). But a simulation of a large blast (right map) shows a high risk of fiery avalanches called pyroclastic flows that reach farther.



COMMENT

EARTH SCIENCE Japanese seismologists reflect on lessons learned **p.146**

CLIMATE Corporations must be made to do much more to cut their emissions **p.149**

FILM How scientists can influence Hollywood **p.150**



OBITUARY Baruch Blumberg, whose work led to hepatitis B vaccine, remembered **p.155**



A WHO official counts mosquitoes and fleas to estimate disease risk in a Rwandan refugee camp.

WHO needs change

The World Health Organization needs major reform to regain its leadership as a convener and provider of scientific and technical knowledge, says **Barry R. Bloom**.

The World Health Organization (WHO) — the United Nations agency created to be the “directing and coordinating authority on international health work”¹ — is increasingly being marginalized and underfunded.

A slow and inadequate response to the recent cholera epidemic in Haiti and the months to years the agency takes to fund projects or approve drugs and vaccines are among the reasons why the WHO has come under attack in recent years for being “ineffective, bureaucratic and political ... and for lacking modern scientific and technical expertise”².

Yet with an explosion of players now involved in global health — from governments and private companies to foundations and non-governmental organizations

(NGOs) — the world urgently needs an organization that can convene the best expertise and provide a centralized resource for health-related knowledge.

The WHO must reinvent itself as this resource. It must re-establish the trust of the international community by improving the transparency of its governance and financing, and by speeding up its responsiveness to countries’ needs.

The WHO has had some towering achievements since it was created in 1948, most notably the Smallpox Eradication Programme. Launched in 1959, this wiped out the disease from the planet by 1977. That success led, in 1974, to the Expanded Programme

for Immunization, which now saves about 2.5 million children’s lives each year from preventable diseases such as measles. Today, the agency has many important, and often unseen, functions¹ — perhaps the most crucial of which is to establish the International Health Regulations. These regulate travel and transport, and as stated in its constitution, give the WHO unique supranational authority “to take all necessary action to ... prevent the international spread of disease.” Yet the WHO is operating in an increasingly complex and fragmented world of global health. Fifty years ago, the main international players were the WHO, the United Nations Children’s Fund (UNICEF), a couple of foundations and the overseas development agencies of some rich countries. Now, thousands ▶

NATURE.COM
For more on
the WHO see:
go.nature.com/cdaljm

C. SATTLBERGER/PANOS

of organizations fund global health (see 'Plethora of players'). As an example, 196 agencies have solicited funds to address the current cholera epidemic in Haiti, each with their own particular stance on how to deal with it³. The entire system is fraught with gaps, redundancies, inefficiencies and enormous burdens on developing countries to sort through the funding maze and be accountable to each player.

Within this maze, the WHO itself is a fragmented organization with a cumbersome governance. The formal governing body is the World Health Assembly, which consists of 193 ministers of health of the member countries. Ministers have a turnover time of around two to three years, which, combined with only once-yearly meetings of the assembly, makes addressing emergencies or long-term problems almost impossible. All member states fall into one of six 'regions', such as the African or South-East Asian regions, each of which is governed by a regional director with great autonomy (the regional offices receive about 75% of the WHO's budget). Regional directors are elected by the countries in the region, and are not appointed by or formally responsible to the WHO's director-general, so some of the agency's activities are uncoordinated and not based on the best scientific evidence.

SLOW TO ACT

Many people who have dealings with the WHO are struck by its bureaucracy — for instance, the number of organizational levels from which countries requesting technical assistance must obtain approval, and the glacial time it takes to get projects funded. The WHO's Green Light Committee, for instance, is supposed to help countries gain access to high-quality drugs to treat people with multi-drug-resistant tuberculosis. It often takes six months for a country's request to be approved, and another year before the first drugs arrive. Meanwhile, thousands of patients are transmitting the disease or dying every day.



Golden years: an immunization programme launched in 1974 was a huge success.

Politicization is also a major problem. The WHO has an executive board consisting of individuals serving in their personal capacity, which was created to prevent countries lobbying for control and to raise the agency's agenda above the level of politics. Sadly, the board — mainly medical doctors appointed by member-country representatives of the World Health Assembly — has itself become highly politicized in regard to major decisions, such as the election of the WHO's leadership. Votes are held in secret and the board is not publicly accountable for many of its decisions. Countries exert huge political and suspected financial pressures on board members to support or oppose certain candidates applying for top positions.

Most worrying of all, the WHO has increasingly failed to demonstrate leadership on the scientific and technical front. In October last year, three days after it was announced that the disease outbreak in Haiti was cholera, the people of the worst-affected region rioted because they believed that the infection had been introduced by Nepalese United Nations peacekeepers. Despite considerable media coverage, WHO headquarters remained silent⁴. When the complete DNA sequences of two Haitian strains were obtained by US university and biotech company researchers, the findings had direct implications for the management of the disease. The Haitian

bacterium turned out to be a South Asian strain that produces a stronger toxin than that made by most cholera strains⁵, making it harder to control. In my view, the WHO should have taken the lead in seeking the source of the Haitian strain, and in ensuring that existing vaccines were made available.

For many countries, the main contributors to disease burden (years of healthy life lost due to premature mortality and disability) are now chronic diseases, including diabetes and cardiovascular or neuropsychiatric conditions, not communicable ones such as cholera⁶. Yet the WHO still has far more expertise in dealing with the latter. In 2008 and 2009, noncommunicable diseases received less than half the funding for communicable diseases; the WHO, with the help of its donors, needs to redress this imbalance.

All these failings have fostered considerable distrust in the donor community — most clearly reflected in donor spending. The WHO budget has two components. The regular budget comprises mandatory contributions by member states (adjusted according to each country's gross domestic product), which in the 2008–2009 biennial budget was less than US\$1 billion of the total \$3.9 billion. Extrabudgetary funds are voluntary contributions by countries or foundations to support specifically targeted programmes. Here, the donors set the priorities (which in itself reflects donors' lack of faith in the agency's ability to prioritize and allocate resources effectively).

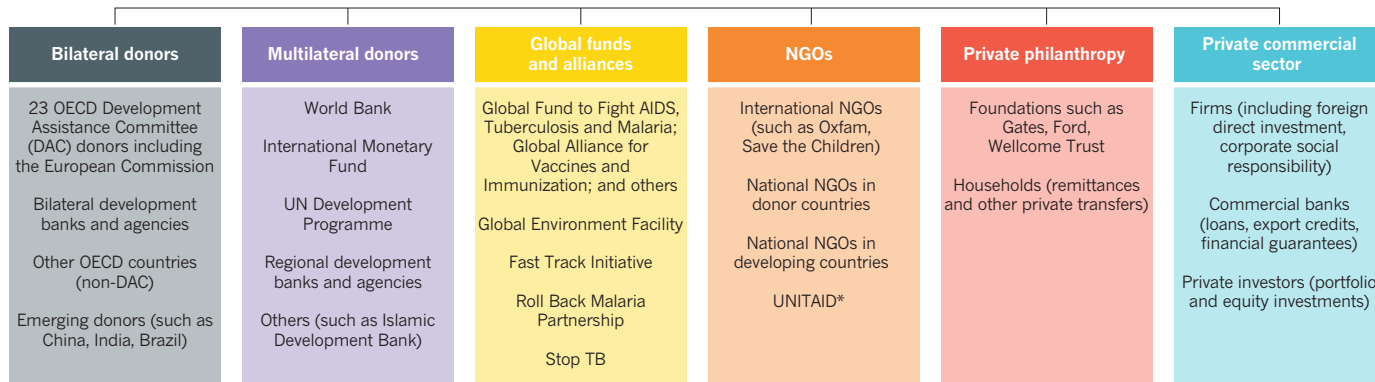
In the past few years, the WHO's biennial budget has fallen by about 10%, and the organization is facing a fiscal crisis⁷. This decline is particularly striking given that the amount of money invested annually in all global health-related activities by donors, including governments and foundations, is estimated to have risen from \$5.6 billion in 1990 to \$26.8 billion in 2010 (ref. 8).

To regain its standing in global health, the WHO needs to win back donor trust by focusing on those areas where it has

N. WHEELER/SGMA/CORBIS

PLETHORA OF PLAYERS

Thousands of organizations now fund global health.



Bilateral donors are agencies of nations that give funds directly to developing-country governments. Multilateral donors are organizations that pool funds from donors and distribute them to many projects in developing countries. Regional development banks lend to governments for local development projects. OECD, Organisation for Economic Co-operation and Development. *UNITAID is a drug-purchase agency funded by several governments.



The WHO has been criticized as unduly bureaucratic and politicized.

an advantage over other foundations, companies and NGOs.

In a remarkably forthright analysis⁹, WHO Director-General Margaret Chan stated last year that “in today’s crowded landscape of public health, leadership is not mandated. It must be earned. And it must be earned through strategic and selective engagement. WHO can no longer aim to direct and coordinate all the activities and policies in multiple sectors that influence public health today.”

A PRESCRIPTION FOR CHANGE

Absolutely, the WHO should focus on the things it can do best. It does not have the budget to be a funding agency like the World Bank or the Global Fund to Fight AIDS, Tuberculosis and Malaria. And it does not have enough ‘on the ground’ staff to be an implementing agency like UNICEF. Instead, it should aim to be the paramount knowledge organization in global health — gathering up the best technical, scientific and practical information and making it accessible to all countries.

Currently, numerous agencies, countries and individuals collect information, for instance on disease burdens, the economics of different interventions and how to monitor and implement particular health programmes. But with the WHO title still carrying significant weight, especially in developing countries, the agency is best placed to provide forums for experts, scientists and health officials worldwide to interact and agree on best practices. The international flu conference held in Washington DC in 2009, and last year’s Global Symposium on Health Systems Research in Montreux, Switzerland, offer valuable models. The WHO should make the results of such exchanges — as well as the data gathered by the health ministries of individual countries and other agencies — available in a central online repository.

Moreover, although the WHO is not

a research institution, it must influence the priorities for research and innovation carried out by academic institutions and industry — in part, to ensure that the concerns of developing countries are addressed. Sadly, research activities at the WHO are dwindling. For the first time in half a century, the Advisory Committee for Health Research failed to meet this year.

Instead of whittling down its advisory effort, the WHO should be persuading the international community to increase its regular budget so that it can recruit the world’s experts in science, technology, health care and health-care economics to evaluate current practices and anticipate possible future needs, such as emerging resistance to antibiotics.

The WHO is also in a stronger position than any other agency to simplify and harmonize the fragmented world of global health. Thousands of donors are now funding projects to address individual diseases. The WHO should foster the integration of health activities, particularly at the local and district level, to ensure patients aren’t siloed into single-disease categories but treated according to whatever complex array of symptoms they display. Mexico recently implemented a programme for mothers and children involving 13 interventions — from immunizations and height, weight and nutritional assessments for children, to breast-cancer and cervical screening for mothers. This is exactly the kind of approach the WHO should be promoting. Its recently launched Alliance for Health Policy and Systems Research initiative, to develop integrated health systems, is an important move in this direction.

Similarly, a few simple steps could transform the tangled web of approval and reporting requirements for countries seeking assistance. For example, the WHO should bring together health ministers and financial experts from developing countries with donors to establish and harmonize common reporting and accountability standards.

To achieve all this, the agency must make three changes to regain the trust of the scientific and donor communities.

First, the WHO should be more inclusive. The agency has had difficulties in creating trusting relationships with civil-society organizations and with industry, particularly the pharmaceutical industry and the private sector. The pharmaceutical industry, for instance, resents the WHO’s essential medicines list, a register of minimum medicine needs for every health-care system, as this stresses the usefulness of inexpensive, off-patent drugs. Yet civil society, NGOs and pharmaceutical organizations are influential actors in global health. They should either have a role in the governance of the WHO (for example, by being rotating members of the executive board) or be more effectively engaged as stakeholders.

Second, the WHO should be more transparent, especially in relation to where funds are spent in the regions. For the executive board that represents the public interest, secret ballots must be abolished and recommendations about leadership and priority-setting made as open and publicly accountable as they are in the UN Security Council.

Finally, the WHO should introduce an external review process. The World Bank has strong internal and external review mechanisms; the Global Fund and the Global Alliance for Vaccines and Immunization both have extensive external review procedures. Not so the WHO. Such a process must engage not just the governing boards but the widest number of stakeholders in a serious review of expectations, performance, priorities and opportunities.

The planet still needs an effective World Health Organization — if a very different one from that created 63 years ago. Organizational transformation is difficult, but just a few key changes could help the WHO to become a farsighted leader, not a lagger, in global health. ■

Barry R. Bloom is Harvard University Distinguished Service professor and Joan and Jack Jacobson professor of public health at Harvard School of Public Health, Boston, Massachusetts.

e-mail: barry_bloom@harvard.edu

1. Lee, K. *The World Health Organization* (Routledge, 2009).
2. Chow, J. C. Is the WHO becoming irrelevant? *Foreign Policy* (8 December 2010).
3. www.disasteraccountability.com
4. McNeil, D. G. Jr Cholera’s Second Fever: An Urge to Blame. *New York Times* (21 November 2010).
5. Chin, C.-S. et al. *New Engl. J. Med.* **364**, 33–42 (2010).
6. Daar, A. S. et al. *Nature* **450**, 494–496 (2007).
7. Sridhar, D. & Gostin, L. O. *J. Am. Med. Assoc.* doi:10.1001/jama.2011.418 (2011).
8. Institute for Health Metrics and Evaluation *Financing Global Health 2010* (IHME, 2010).
9. WHO *The Future of Financing for WHO* (WHO, 2010).



KYODO/REUTERS

Rescue workers search for survivors in the village of Nodamura in Iwate Prefecture, northern Japan.

Rebuilding seismology

Two months on from the earthquake and tsunami that hit their country on 11 March, five Japanese seismologists reflect on what they have learned from it so far.

TAKESHI SAGIYA Integrate all available data

Nagoya University, Japan

If historical records had been more complete, and if discrepancies between data had been picked up, we might have been alert to the danger of a magnitude-9 earthquake hitting Tohoku, even though such an event was not foreseen by the Japanese government.

In 2002, the Headquarters for Earthquake Research Promotion of the Japanese government released a long-term evaluation of the likelihood of subduction-zone earthquakes in the Tohoku region. It estimated an 80–90% probability that the area would have

a large earthquake of magnitude 7.7–8.2 in the next 30 years. But the probability of a magnitude-9 earthquake affecting a 400–500-kilometre area was not specifically mentioned. As a member of the working group involved in the evaluation, it is with great regret that I reflect on the causes of this failure.

The long-term evaluation was based on the statistical analysis of the complete historical earthquake record for the past 400 years. However, the Tohoku earthquake clearly shows that 400 years is too short a time period to evaluate seismic activity. In fact, during the past five years, geologists of the Geological Survey of Japan have reported that a great tsunami inundated the coast of the Sendai area in AD 869. It was probably of comparable size to the tsunami that hit after the March earthquake. Unfortunately, those findings came too late to be considered

in the evaluation of, and countermeasures against, tsunamis.

One important lesson from this experience is that if we take an empirical approach to evaluating or forecasting natural disasters, all the available information should be taken into account — even though some records have large uncertainties — and all the possibilities should be considered, regardless of their likelihood.

Another weakness of the long-term evaluation was clear discrepancies between the different observational data that went into it. In the past decade, Global Positioning System (GPS) investigations have indicated that the plate boundary along the Japan trench is almost fully locked



➔ WWW.NATURE.COM/JAPANQUAKE

in place and is not sliding^{1,2}. But it has been recognized for long time that, in the Japan trench, the ratio of cumulative fault slip of large earthquakes to plate motion (the seismic coupling coefficient) is only about 30%. There has been no explanation of how the remaining 70% of plate motion is accommodated.

Japanese seismologists had noticed this discrepancy, but we had not seriously considered its potentially disastrous implication. In 2001, Ichiro Kawasaki of Kyoto University suggested that a significant portion of the plate motion is accommodated as 'afterslip' — further slippage after an earthquake has taken place — and other aseismic faulting behaviour in the northern part of the Japan Trench³. But the discrepancy was not resolved for the southern part, the main source region of the magnitude-9 earthquake.

Thus the second lesson of the great Tohoku earthquake is that we should not overlook inconsistent data, but instead strive to integrate observational information with different temporal and spatial scales.

Earth science is multidisciplinary. The Japanese seismology community now needs to review all the seismic, geodetic, geomorphological and geological data to find information missing from the current evaluation, and to resolve any inconsistencies.

HIROO KANAMORI

Prepare for the unexpected

*California Institute of Technology,
Pasadena, California*

The 2011 Tohoku earthquake caught most seismologists by surprise because no earthquake of a magnitude greater than 8.5 was known to have occurred in this region. Extensive analyses of seismic, GPS and tsunami data conducted since the earthquake make it clear that unusually large strain and stress release occurred in a relatively narrow zone within 150 kilometres of the Japan trench (see 'Pressure zone'). The amount of strain release is nearly an order of magnitude larger than what we have seen in other mega-thrust earthquakes. The strain must have accumulated in this zone for nearly 1,000 years, with the plate convergence rate of about 9 centimetres per year. Finally, the stress exceeded the local strength of rock and failed, causing the magnitude-9 earthquake.

Despite the extensive GPS network in Japan, this localized large strain build-up had not been detected because the area is 200 kilometres offshore. An important lesson we have learned is that such a large strain can accumulate in the shallow plate boundary (see 'Pressure zone') — it had

been considered able to accommodate only 10–20% of the strain released during the 2011 event.

Thus, for monitoring a strain build-up large enough to eventually cause a magnitude-9 event, it is important to accelerate the effort to develop ocean-floor GPS technology, and to promote research to detect unusual seismogenic structures responsible for such large strain build-up (for example, rough plate interfaces).

Even if we understand how such a big earthquake can happen, because of the nature of the process involved we cannot make definitive statements about when it will happen, or how large it could be. It will be a very rare event but, once it happens, it will have grave consequences. So we must try our best to be prepared for the unexpected. Building a robust infrastructure is most important. However, there is a limit to what we can do with the available technology, so we need to seriously consider the acceptable trade-off between benefit and risk.

YUJI YAGI

Enhance ocean-floor observation

University of Tsukuba, Japan

The process by which the Pacific plate is subducting beneath the Japanese archipelago is not smooth. Strain can accumulate where parts of the plate are unable to slip because they are stuck to the overlying continental plate. Major inter-plate earthquakes release the resulting accumulated strain, allowing slip to resume. The 2011 Tohoku earthquake released a huge slip deficit, which was revealed at least in 2004 (ref. 1), or even as

far back as 2000 (ref. 4) by the GPS network, and anticipated from a discrepancy between geological and geodetic rates of crustal deformation by Yasutaka Ikeda of the University of Tokyo in 1996 (ref. 5).

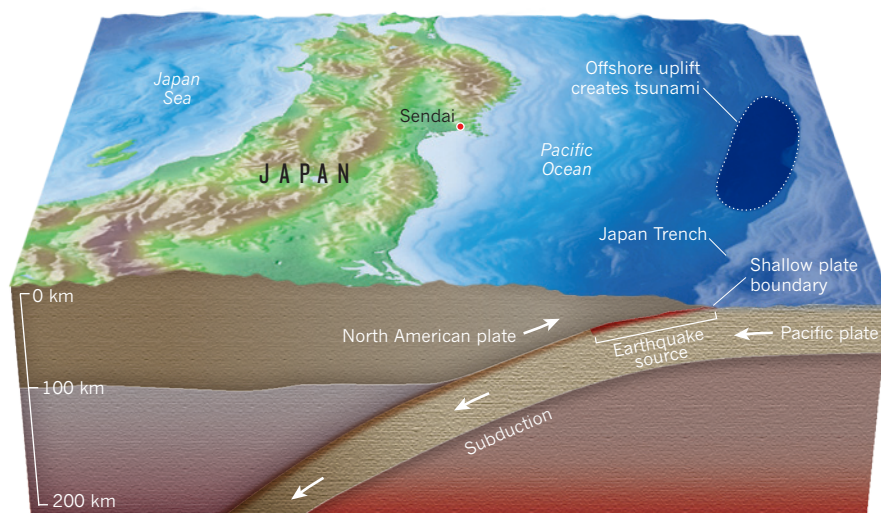
The distribution of areas where the plates are stuck (known as asperities) is usually inferred from the analysis of seismic event data. And, on the basis of this 'characteristic earthquake model', the regions likely to experience future earthquakes are estimated by responsible commissions in Japan such as the Headquarters for Earthquake Research Promotion. Most seismologists in Japan assume reasonably that areas beyond the asperities are free from earthquakes for long periods of time because the strain there is released by aseismic slip. We need to confirm this assumption.

Great earthquakes occur not only as a result of the combined effect of simultaneous strain release at several assumed asperities. The accumulation of strain suggested by crustal deformation or GPS observation has not been adequately considered in the government's seismic-hazard estimations in Japan because of the low resolution of the data. This oversight reduces the total amount of assumed slip deficit. This is why most seismologists did not recognize the risk of a huge Tohoku earthquake.

We are not able to obtain accurate information about the patterns of great earthquake occurrences in time and space because of the long period of the seismic cycle. Therefore, we cannot assess the probability of future great earthquakes on the basis of catalogues of recent seismicity alone. Great earthquakes are exceptional events reflecting very long-term deformation processes in a subduction zone. To develop a comprehensive understanding of them we must also consider geodetic, geological and geomorphological information about crustal deformation. ▶

PRESSURE ZONE

Strain accumulates in the Japan Trench when the Pacific plate gets stuck while sliding beneath Japan.



► In general, for predicting long-term seismicity, methods must be developed to extract information about elastic and anelastic strain from observational data. Pressingly, ocean-floor observation must be enhanced to allow high-resolution estimates of slip deficit distributions along the plate boundaries.

MASUMI YAMADA

Warnings work, but must be better

Disaster Prevention Research Institute, Kyoto University

The Japan Meteorological Agency has one of the most advanced systems in the world for providing real-time warnings of tsunamis and earthquake shaking. The earthquake early warning system, which provides information about strong shaking within seconds of a quake, has been in place since 2007 and has provided more than 10 warnings of strong earthquakes — by cellular phone, television, radio and local-community speaker system. But it could be better.

The system detected the earthquake off the Pacific coast of Tohoku and, about 8 seconds after the first primary wave arrived at the closest seismic station, issued a warning to the public in the region close to the epicentre. Twenty seven bullet trains were stopped without derailments in this region. Three minutes later, warnings for very large tsunamis were issued to Iwate, Miyagi and Fukushima prefectures. The damaging waves arrived 15–20 minutes later at the closest shores.

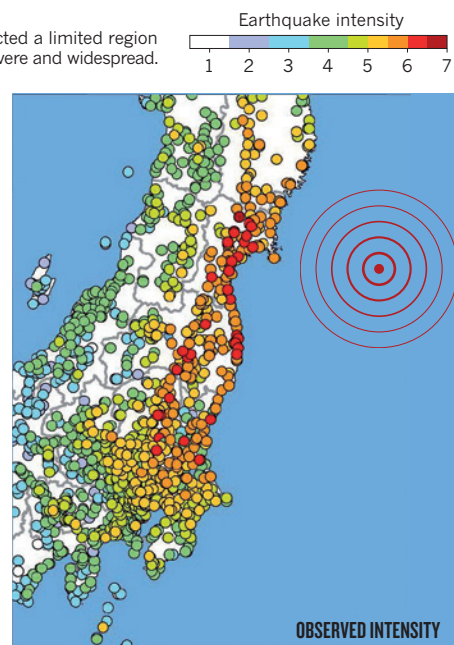
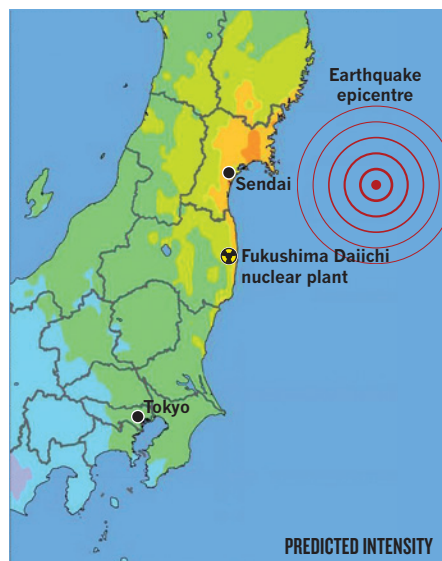
However, the overall performance of the system was not satisfactory, mainly because of the complex character and relatively small amplitude of the beginning of the rupture. The system underestimated ground motion and tsunami heights, so the large population in the greater Tokyo region, where many areas experienced strong and damaging shaking, received no warning (see 'False comfort'). That said, updates did improve as more information became available.

Early warnings for strong shaking were broadcast more than 70 times for aftershocks. The system worked well for these smaller events, but there were some errors in determining event locations because of the complication of simultaneously occurring earthquakes.

The unexpected character of the seismic data at the start of the earthquake fooled the early warning system's algorithms. But the system has the potential to work well for the next great earthquake — such as the widely expected Nankai earthquake in the Kansai region — if technical improvements are

FALSE COMFORT

A warning system based on initial seismic signals predicted a limited region of intense shaking. The actual shaking was much more severe and widespread.



SOURCE: M. YAMADA & JAPAN MET. AGENCY

made to recognize great earthquakes quickly. The earthquake early warning system in Japan should become a truly effective mitigation tool in a society that has already accepted and learned to expect such information.

JIM MORI

Design buildings for greater shakes

Disaster Prevention Research Institute, Kyoto University

The Tohoku earthquake came as a frightening and disheartening surprise to Japanese seismologists, who thought they could roughly predict the locations and sizes of plate-boundary earthquakes along the subduction zones of Japan. Some 400 years of historic records have proved to cover too short a time period to be a reliable guide to the occurrence of the largest earthquakes, even for this very seismically active region. With hindsight, we must now re-examine the Japanese geological data for events of a similar magnitude.

Usually in seismology, the rates at which future earthquakes are expected to occur have been largely derived from the statistics of repeating events on faults.

However, the 2011 earthquake shows that rarer and much larger earthquakes can also occur in the same fault zone as smaller events. So past statistics are not always sufficient. One way to improve our understanding is with efforts to measure the local accumulation of stress near faults and to estimate the absolute stress and strain levels at which earthquakes happen.

For example, stress can be measured directly within boreholes; frictional strength can be inferred from temperature measurements after large earthquakes; and directions of regional stress fields can be determined from patterns of ground displacements recorded by GPS arrays.

Using such techniques, we may be able to determine how close a fault is to failure, and thus estimate the earthquake risk more directly. The size of an impending event will still be difficult to determine, but maximum sizes might be estimated from the stress measurements. The regional deformation of eastern Japan is now being studied to try to discern what stress and strain the area underwent before the 2011 quake.

The evaluation of strong shaking from great earthquakes is a related issue of great importance — for building safety standards around the world — but a long record of accurate ground motions from past earthquakes is lacking. As measurements have improved over the past few decades, our expectations of the severity of the motions have steadily increased. It is unlikely that we have seen the worst. Greater shaking, especially at long periods of several seconds, is a possibility that must be considered in hazard planning. Such information is needed for the safe design of tall buildings in our modern cities. ■

1. Nishimura, T. *et al. Geophys. J. Int.* **157**, 901–916 (2004).
2. Hashimoto, C., Noda, A., Sagiya, T. & Matsu'ura, M. *Nature Geosci.* **2**, 141–144 (2009).
3. Kawasaki, I., Asai, Y. & Tamura, Y. *Tectonophysics* **330**, 267–283 (2001).
4. Nishimura, T. *Spatiotemporal Change of Interplate Coupling in Northeastern Japan Inferred from GPS Data* [in Japanese]. PhD thesis, Tohoku Univ. (2000); available at <http://go.nature.com/7ikqir>
5. Ikeda, Y. *Active Fault Res.* **15**, 93–99 (1996).



Floating houses in the Netherlands remain a small experiment as scaling requires regulations and improvements to business models.

CLIMATE ECONOMICS

Corporate greening falls short

Gail Whiteman is unconvinced by an argument that naked greed and market forces will drive businesses to cut their emissions.

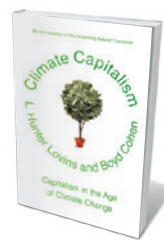
Sustainability is no longer a fringe topic. Corporations are routinely taking steps to reduce their carbon footprints and investing in green business measures. Yet important ecological thresholds are still being breached. Companies are the main engines of socioeconomic change, and many are actively creating market advantages from sustainability. Why then are emission cuts still so abysmally small?

Instead of sorting through this dilemma, *Climate Capitalism* is filled with positive stories of how profit-seeking firms are helping save the planet. Through engaging examples, the book seeks to demonstrate that capitalism and “naked greed” are “powerful motivations for solving the problem of climate change”. Both authors have first-hand experience of helping organizations to improve their green credentials: Hunter Lovins’ consultancy

advises the United Nations and the US retail giant Walmart on sustainability, and Boyd Cohen, a former professor of sustainable

entrepreneurship is chief executive of a carbon-management company. But their book is not a critical analysis.

Many business leaders recognize that economic forces are insufficient for resolving the complex challenges that face Earth systems. Chief executive of the global transportation and logistics company TNT, Peter Bakker, has told me: “As a company we can reduce our carbon footprint dramatically. But the world’s still driving off a cliff. We need a system change.” Senior business consultants say that they need more than good examples. They want to know how they can overcome barriers to change inside and outside a firm and how to let companies know if what they are doing is having a significant



Climate Capitalism: Capitalism in the Age of Climate Change

L. HUNTER LOVINS
AND BOYD COHEN
Hill & Wang: 2011.
400 pp. \$27.95

effect on the world’s resources. How should companies judge their impacts on ocean acidity, for example? Or on atmospheric pollutants other than carbon dioxide?

Science can help. Cutting emissions is essential, but Earth’s natural processes are multi-dimensional and interlinked, making it difficult to measure the specific effect of emissions cuts. A useful tool is the ‘planetary boundary’ concept of Swedish environmental scientist Johan Rockström and his colleagues. In 2009, they derived limits to seven key Earth systems — such as climate change, ocean acidity, stratospheric ozone levels — that, if breached, would be detrimental to human survival (J. Rockström *et al. Nature* **461**, 472–475; 2009). Lovins and Cohen nod to this work, but don’t discuss how to bring it into the boardroom.

The case studies in *Climate Capitalism* are appealing, but the descriptions are based more on media reports than on research, so some projects are exaggerated. For example, the book describes the ‘floating houses’ piloted by the Dutch company Dura Vermeer as “popular with locals in the

➔ **NATURE.COM**
A review of Nicholas Stern’s *A Blueprint for a Safer Planet*.
go.nature.com/ppw80g

Netherlands and foreigners, alike". In fact, they remain a small-scale experiment mainly in one city, Maasbommel. Scaling them up to floating communities has proved difficult owing to a lack of regulations and good business models, problems not discussed in the book.

The most regrettable dearth of analysis is in the chapter on carbon markets. The authors argue that the market will work wonders, if we let it. They acknowledge that some carbon-offset schemes are scams and that the European Union's Emission Trading System (EU ETS) had problems at the beginning, then quickly go on to call for carbon markets to be rolled out widely. There is no alternative, Lovins and Cohen say, because a carbon tax is not appealing to US politicians.

Not so. Economic studies suggest that a carbon tax may be more cost-efficient than a cap-and-trade system (B. B. F. Wittneben *Energy Policy* 37, 2462–2464; 2009). It provides a clear price signal, requires less bureaucracy, which reduces costs, and accrues revenue straight to the government that collects it. Most importantly with such an approach, there is no upper limit on emissions reduction. So why rule out a carbon tax as one of the tools of climate capitalism? Carbon taxes have indeed been successful elsewhere, for example, in British Columbia, Canada.

Nor are the market-based solutions as great as *Climate Capitalism* makes out. As a result of intense lobbying, some big polluters, particularly in the utilities sector, made windfall profits from the EU ETS without reducing emissions. Carbon-market fraud is also on the rise. One estimate suggests that such crime has cost the EU at least €5 billion (US\$7.3 billion) since 2007 (M. K. Dorsey and J. Whittington *Carbon Market Europe* 9, 7; 2010). In January, the Commission suspended EU ETS transactions for a week after a series of frauds.

Naked greed is not a motivating force for sustainability; it is a disruptive one. Although the positive examples that Lovins and Cohen present are inspiring, they are not enough. Binding international agreements and new governance mechanisms that address our integrated environmental issues — including, but not restricted to, emissions cuts — will compel businesses to engage more deeply with science. ■

Gail Whiteman is a professor in the Department of Business Society Management, Rotterdam School of Management, Erasmus University, Rotterdam, the Netherlands.
e-mail: gwhiteman@rsm.nl



Sigourney Weaver (front right) in *Avatar*: director James Cameron sought scientific advice over her lines.

COMMUNICATION

Popcorn and Petri dishes

Cinemas are today's scientific lecture halls, finds **Kevin Hand** in a book probing how research enriches film.

In the autumn of 2007 I was at the American Geophysical Union conference in San Francisco, California, when my mobile phone rang. The number was blocked; as a recently minted PhD with credit-card debt, I assumed it was a company trying to track me down. Nevertheless, I answered.

To my surprise, the voice at the other end asked, "Dr Hand? I've got James Cameron on the line from New Zealand. Do you have a few minutes to help him?" I had worked with the director before and knew he was working on his next film. Cameron got on the line and described to me life on his distant planet. He needed to feed Sigourney Weaver a few lines of science jargon. Within 20 minutes we had it figured out. I put my mobile away and went on with the meeting.

Later that week I would give my talk about Jupiter's moon Europa to some 200 attendees. Two years later, Cameron's film *Avatar* reached many tens of millions of viewers.

Granted, should we some day discover life on Europa, it will dwarf the impact of any Hollywood movie. But we won't get a chance to search for that life unless the tax-paying millions that watch films care enough about our potential discoveries to invest their dollars in them. From supercolliders to NASA missions, big science is going extinct, in part because politicians and public are missing the relevance; the story isn't there. Science stands to benefit from a symbiotic relationship with those who know how to tell stories, notably film-makers.

In *Lab Coats in Hollywood*, scholar of science communication David Kirby analyses the interplay between science consultants and Hollywood film-makers. His approach is academic, but numerous examples and interesting historical details make for an enjoyable read. The book is in part a handbook for those who might try to influence the way in which science and technology are portrayed on the big screen.

Kirby notes that science advisers often labour under the misconception that there is a tension between accuracy and entertainment in film-making. On the contrary: for film-makers, he says, "there is only entertainment". The best approach for a scientist is thus to focus on how accurate science can make the story better. Rare directors such as Stanley Kubrick and Cameron are obsessed with scientific integrity, but the majority of them use it as ornament.

Although some of Kirby's conclusions are obvious, I found two concepts particularly intriguing. He argues that films are a powerful "virtual witnessing technology", in that they provide the mechanism for people to gain



Lab Coats in Hollywood: Science, Scientists, and Cinema
DAVID A. KIRBY
MIT Press: 2011.
264 pp. \$27.95

access to evidence for scientific knowledge. He borrows from the late US evolutionary biologist Stephen Jay Gould, who argued that the “need for direct witnessing is what separates scientific practice from religious faith”.

So films serve as the descendant of the theatrical public science lecture, such as those implemented in the nineteenth century by Britain's Royal Institution, in which luminaries like Michael Faraday demonstrated experiments to a live audience. The public could verify the outcomes with their own eyes. Cinema can have a similar effect, even when the viewer is witnessing a simulated experiment on the screen.

Kirby also explores the concept of the “diegetic prototype” — inventions demonstrated in the fictional world that catalyse developments in the real world. For example, the imagined prototypes for rockets and lunar exploration presented in Fritz Lang's 1929 film *Woman in the Moon* and the 1950 film *Destination Moon*, whose technical adviser was science-fiction writer Robert A. Heinlein, helped to lay the foundation for the political, public and scientific acceptance needed to move space endeavours forward.

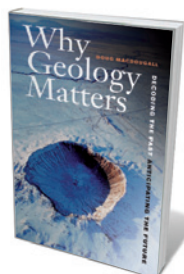
Yet, as most researchers are aware, science regularly suffers at the hands of a good story. From the extremes of weather to vaccines and autism, anecdote frequently trumps data. The response of the scientific community is often to seek new and better data. But on the landscape of knowledge, what society needs is not a better map, but a better description of how to navigate the terrain. Science is best conveyed to the public in a compelling narrative.

Love it or hate it, Hollywood remains influential. We may, as scientists, be disappointed that society prefers character over content. As Carl Sagan remarked in his 1996 book, *The Demon-Haunted World*: “if, for whatever reason, people dislike the stereotypical scientist, they are less likely to support science”. But a good on-screen portrayal of science and scientists allows us to incorporate valuable facts and ideas into entertaining stories. And as Kirby details, new mechanisms for collaboration are emerging. Over the past three years, the National Academy of Sciences' Science and Entertainment Exchange (go.nature.com/pcxgsm) has connected film-makers with scientists to brainstorm about how to use good science to build a better story.

Lab Coats in Hollywood provides a framework for scientists to better understand how to influence good storytelling with accurate information. We don't all wear lab coats, but a few more beakers behind the scenes could go a long way towards enhancing critical thinking in modern society. ■

Kevin Hand is a planetary scientist at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA.
e-mail: kevin.p.hand@jpl.nasa.gov

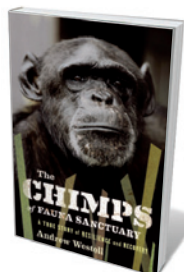
Books in brief



Why Geology Matters: Decoding the Past, Anticipating the Future

Doug Macdougall UNIVERSITY OF CALIFORNIA PRESS 304 pp. \$29.95 (2011)

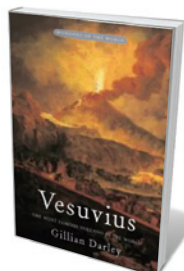
From mountain ranges to meteorite craters, the story of our planet is embedded in its rocks. In a wide-ranging and entertaining overview of the field, geologist Doug Macdougall explains how Earth scientists have unravelled the secrets of geological time, plate tectonics, volcanoes, earthquakes, past climates and the fossil record. He also muses on the role of geology in addressing pressing problems such as climate change and the continued provision of energy.



The Chimps of Fauna Sanctuary: A True Story of Resilience and Recovery

Andrew Westoll HOUGHTON MIFFLIN HARCOURT 288 pp. \$25 (2011)

Biologist-turned-writer Andrew Westoll relates his experience as a volunteer at Gloria Grow's chimpanzee sanctuary near Montreal in Canada. In a vivid narrative, he describes how he got to know the rescued chimps. Many of them had come from a biomedical facility or were kept as pets, and some showed psychological problems as a result. He tells how the more time he spent with the animals, the more he learned about their behaviour and the stresses they endured in their captive lives.



Vesuvius: The Most Famous Volcano in The World

Gillian Darley PROFILE BOOKS 224 pp. £15.99 (2011)

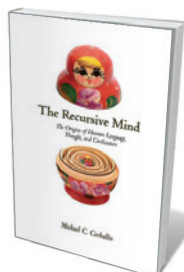
Looming over the bay of Naples, Vesuvius is the only active volcano on the European mainland. It has mesmerized locals and visitors alike for two millennia with its majestic beauty, geological treasure trove and vicious eruptions — although none has since matched the devastation of the explosion that destroyed Pompeii in AD 79. In an elegant exploration of the volcano's allure, author Gillian Darley recounts how Vesuvius has inspired literature, imagery and science, from the first serious studies of volcanic activity during the Enlightenment to paintings by Andy Warhol.



Listed: Dispatches from America's Endangered Species Act

Joe Roman HARVARD UNIVERSITY PRESS 368 pp. \$27.95 (2011)

In his tour of the places where rare species have stood in the way of dams and developments across the United States, ecologist and writer Joe Roman analyses the impact of the 1973 US Endangered Species Act. Using examples such as the whooping crane, North Atlantic right whale and the purple bankclimber — a freshwater mussel enmeshed in a water war with the city of Atlanta, Georgia — he argues that protecting biodiversity benefits economies and well-being alike, showing that species extinctions have a tangible impact on humans.



The Recursive Mind: The Origins of Human Language, Thought, and Civilization

Michael C. Corballis PRINCETON UNIVERSITY PRESS 288 pp. \$29.95 (2011)

What makes us human? Psychologist Michael Corballis argues that it is our ability to embed thoughts in other thoughts — known as recursion — rather than language that allows us to conceive of ourselves and others, and to witness the passing of time. He suggests that recursive minds were crucial to the survival of our Pleistocene ancestors and led to the emergence of societies, toolmaking and culture as well as communication.



Tania Kovats's *TREE* (2009), installed at London's Natural History Museum to mark Charles Darwin's bicentenary.

ART

Taking the long view of nature

A monograph highlights how artist Tania Kovats views geological and evolutionary time, notes **Colin Martin**.

To celebrate Charles Darwin's bicentenary in 2009, a 70-metre longitudinal section of an oak tree was embedded in the ceiling of London's Natural History Museum. *TREE* (2009) is the work of British artist Tania Kovats. Two decades of her reflections on geological structures and evolutionary change are now celebrated in an illustrated monograph.

The passage of time is a constant theme. Many of Kovats's installations mimic geological processes, including erosion, eruption, compression and subsidence. In her replica mountain-making machine *Mountain* (2001), layers of black, white and grey wax buckle and fold as they are compressed by a piston.

The apparatus is based on one used by the nineteenth-century American engineer Bailey Willis to demonstrate how the Appalachian Mountains were formed. Other sculptures

shown in the book evoke the basalt columns of the Giant's Causeway in Northern Ireland, the slashed geometries of strike-slip faulting, the structures of minerals and geological maps.

Drawing is central to Kovats's practice, and some of her fine preparatory pencil sketches of geological works, tree rings and silhouettes are reproduced. Kovats observes the natural world as accurately as any scientist, but her art is more than a direct record. For example, *The Museum of the White Horse* (2007) — inspired by the Uffington White Horse, a giant equine form cut into Oxfordshire chalk downland by Bronze Age artists — assembles drawings and horse-related paraphernalia in a gaily decorated



Tania Kovats
JEREMY MILLAR AND
PHILIP HOARE
Lund Humphries:
2011. 144 pp.
£35, \$70

horsebox to comment on humanity's varied relationship with the horse throughout history.

In December 2009, Kovats travelled in Darwin's footsteps to the Galapagos Islands to observe geological and biological morphologies on vastly different scales — from volcanoes to barnacles. The drawings that resulted include her sketch of the battered copy of Darwin's *Voyage of the Beagle* that she brought with her. Work by Kovats, and another 11 artists who have undertaken Gulbenkian residencies on the Galapagos Islands, will tour Edinburgh, Liverpool and Lisbon next year.

Kovats's interest in the changing landscape fuels her latest artwork. This summer she plans to launch a wet-meadow raft, complete with waterbird nests, along a London canal to provide hassled city dwellers with a restorative glimpse of nature. ■

Colin Martin is a writer based in London.
e-mail: cmpubrel@aol.com

➔ **NATURE.COM**
The earliest sketch
of evolution?
go.nature.com/lkdqhn

ART

Pinball wizardry

A European show reveals new ways of thinking about energy, **Daniel Cressey** learns.

What is energy? For the past two years, a group of 27 European physicists, architects and artists have been working together to build installations that illuminate different aspects of the concept. Developed initially for the 2010 Venice Architecture Biennale, the fruits of their labours are on show this month at London's Architectural Association (AA) School of Architecture.

The exhibits in *Beyond Entropy* include an impossible pinball machine, with hundreds of balls in play inside a clear box, their energy just out of reach once they shoot behind a mirrored divider. Another recreates a 'time machine', proposed in 1899 by French absurdist writer Alfred Jarry, from an array of giant wooden flywheels. Holograms and images of the Moon's cratered surface also show how energy may be embodied in pictures.

These collaborative pieces are the brainchild of Stefano Rabolli Pansera, an Italian architect who teaches at the AA school. After worrying about the disappointingly conventional approach to energy that most designers take, Pansera sought new ways of thinking from other disciplines. He hopes that the show will inspire fresh approaches from architects — and that it will also set the minds of those who are interested in physical concepts spinning in new directions.

The project aims to consider energy as a broad concept, related to the idea of continuous transformation. Eight small teams



A prototype 'time machine' based on an 1899 idea explores the theme of mechanical energy.

of researchers, each including an artist, an architect and a scientist, explored energies electrical, mechanical, potential, mass, sound, thermal, chemical and gravitational. The groups visited CERN — Europe's particle-physics lab near Geneva, Switzerland — to see the enormous amounts of energy being consumed there, and developed their ideas in workshops.

Andrew Jaffe, a cosmologist at Imperial College London, worked on the time machine. Tasked with exploring mechanical energy, his team discussed Jarry's early expression of time as a fourth dimension in his 1899 pseudoscience treatise *How to Construct a Time Machine*. They built a version of his fantasy mechanism, consisting of "three rapidly rotating gyrostats with shafts parallel to the three dimensions of space".

"I found it fascinating that Jarry wrote his piece around the same time as H. G. Wells's *The Time Machine* and well before Einstein. The idea of time as a fourth dimension must have been bouncing around in the popular

culture", says Jaffe.

Roberto Trotta, also a cosmologist at Imperial College London, who worked on the pinball machine developed by the potential-energy team, said he welcomed the fact that the cross-disciplinary conversations went beyond the normal 'top-down' popular lectures that he regularly gives. "I also want to explore a more level approach, where there is an ongoing reflection and dialogue," he explains, "where the scientist doesn't just show up and tell people how nature actually works."

How some of these weird and wonderful creations relate to energy is not always obvious at first glance. But this could be part of the group's plan. "We will be happy if people come away confused," says Trotta. "Not in a bad way, but rather, inspired to think of energy from a different perspective." ■

Daniel Cressey is a reporter at Nature.



'Impossible pinball' exploits potential energy.

CORRESPONDENCE

Collaborations span 1,553 kilometres

Global communications are making physical distances between research collaborators ever more irrelevant. Our analysis of 39 million authors' addresses in research publications indexed in Thomson Reuters' Web of Science for 1980–2009 reveals some surprising exceptions to this trend.

We found that collaboration in the natural sciences over this period spanned the longest distances, with the medical and life sciences catching up fast. Collaboration was most confined for the social sciences and humanities, but this too has been extending rapidly over the past decade. Overall, the average collaboration distance increased more or less linearly from 334 kilometres in 1980 to 1,553 km in 2009.

Countries in the Southern Hemisphere, such as New Zealand and Australia, and several developing countries in the tropics contributed to some of the higher scores. Turkey, Iran, India and most eastern European countries are among the least globalized in terms of long-distance scientific collaboration. Surprisingly, some of the countries with rapid growth in scientific-publication output, such as China and Brazil, show hardly any increase in globalization, or even a slight decrease.

Researchers are in a better position than ever before to engage in long-distance collaborations. This, combined with the increasingly dispersed location of research centres across the globe, could account for the 5.4% annual growth in collaboration distances.

Robert J. W. Tijssen, Ludo Waltman and Nees Jan van Eck
Centre for Science and Technology Studies (CWTS), Leiden

*University, the Netherlands.
tijssen@cwts.leidenuniv.nl*

Open up monitoring of deep-sea drilling

We need independent monitoring of the hydrocarbon industry's deep-sea activities to widen understanding of their potential impact (*Nature* **472**, 152–154; 2011).

The facilities for this monitoring already exist. Real-time observatory systems that deliver data from the sea floor and water column over the Internet can be combined with sample collection and impact assessments of the marine industry to help untangle natural from anthropogenic changes and aid sustainable use of resources.

Efforts such as the European Seas Observatory NETwork, the European Multidisciplinary Seafloor Observatory and Ocean Networks Canada are enabling remote observing, early warning, data discovery and archiving. Researchers are also working with industry, for example on the Deep-ocean Environmental Long-term Observatory System project.

With the United Nations, the intergovernmental Group on Earth Observations — the body coordinating the set up of a Global Earth Observation System of Systems (GEOSS) — could advance these initiatives and disseminate agreed terms to stakeholders.

Public availability of such collective data would also aid mitigation and make an important contribution to GEOSS.

Henry A. Ruhl *National Oceanography Centre, Southampton, UK.
h.ruhl@noc.ac.uk*

Imants G. Priede *Oceanlab, University of Aberdeen, UK.
Competing financial interests: declared (see <http://dx.doi.org/10.1038/473154b> for details).*

Modernize Ukraine's university system

Twenty years after independence, the Ukrainian government is trying to stay competitive in scientific areas such as aerospace, applied mathematics, theoretical physics, energy and organic farming. But its higher education is still tied to the old Soviet system.

The government's intended reforms do not go far enough towards meeting international standards. For example, Ukrainian scientists are trained for a Soviet-style Doctor of Sciences (DSc) degree, which is not based on original research or external peer review. Scholars instead spend 10–20 years on unproductive, essentially bureaucratic work. Therefore, despite a doubling in the number of DSc students over the past 20 years (see go.nature.com/f8agxb), international ratings for Ukrainian universities have remained low.

A paucity of publications in international peer-reviewed journals also stems from Ukraine's academic promotion system, which fosters inertia among research scientists, and from poorly developed skills in foreign languages.

Ukraine's universities need to adopt internationally recognized standards, promote autonomy under democratic and competent management, and support academics to encourage them to stay at home.

Alexander Gorobets *Sebastopol National Technical University, Ukraine.
alex-gorobets@mail.ru*

Tracking China's publication boom

The 2010 Asia-Pacific Publishing Index of the Nature Publishing Group (NPG) family of

journals (go.nature.com/fqm8s3) analyses publication growth and research trends in this important and dynamic region. But omission of a crucial component of the analysis may have distorted its performance comparisons between countries.

The documented increase in NPG publications in the region for 2000–10 does not take into account the increase over that period in the number of articles per journal per year (see Scopus bibliographic database; www.scopus.com) or the rise in the number of NPG journals (from 7 to 17).

Comparing the percentage publication share of Japan and China in NPG journals between 2000–10 reveals that Japan's barely changed (increasing from 7.5% to about 9.5%). China's grew much faster (from 0.6% to 5%) over the period, indicative of a vibrant and expanding research community.

This picture is very different from the results presented in the supplement.

Thomas S. Jones, Andrew M. Plume *Elsevier, Kidlington, Oxford, UK.
a.plume@elsevier.com*

Publisher's note: *We noted in the supplement's introduction that removing the effects of the new publication Nature Communications reduced the growth in papers from the Asia-Pacific region from 25.4% to 14.6%. We stand by our conclusion that growth in the number of NPG papers over that period for Japan was modest and spectacular for China.*

CONTRIBUTIONS

Correspondence may be sent to correspondence@nature.com after consulting the guidelines at go.nature.com/cmchno. Readers are welcome to comment online at www.nature.com/nature.

Baruch Blumberg

(1925–2011)

Geneticist whose discovery led to a vaccine for hepatitis B.

In the 2,000 years since Hippocrates described the skin-yellowing condition he termed *ikterus*, no single event has been more pivotal to the understanding and prevention of viral hepatitis than the discovery of the 'Australia antigen' by Baruch (Barry) Blumberg. This antigen, which Blumberg found in the blood of an Australian Aborigine, turned out to be the surface antigen of hepatitis B virus. The finding ultimately led to a test to screen blood donors for the virus and to a hepatitis vaccine.

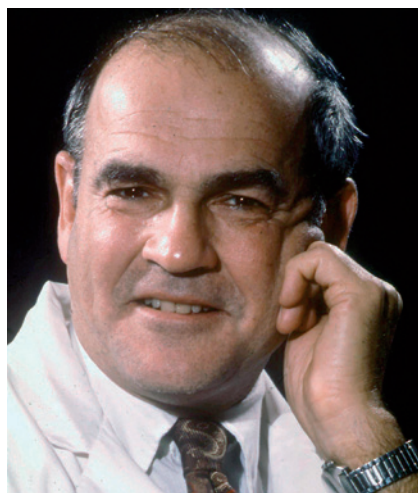
Although the outcome of Blumberg's research was monumental, his findings were initially serendipitous. In the late 1950s, while working at the US National Institutes of Health (NIH) in Bethesda, Maryland, he showed that cholesterol-transporting proteins called β -lipoproteins exist in varied forms in different populations as a result of genetic polymorphisms. In 1963, while heading the new Geographic Medicine and Genetics Section of the NIH, Blumberg and his co-workers unexpectedly found a precipitate in agar that did not take up the blue stain for lipids but stained red for protein. Blumberg had been using serum samples from haemophiliacs who had received multiple blood transfusions to provide antibodies that would react in agar with diverse proteins. The initially designated 'red antigen' was the result of an antibody in the serum of a haemophiliac from New York interacting with an antigen in the serum of an Australian Aborigine, and was later named the Australia antigen. Others might have dismissed this obscure finding as an irrelevant curio. But Blumberg's hypothesis-generating mind was set in motion.

After finding that the antigen was 100 times more prevalent in patients with leukaemia than in healthy blood donors, Blumberg proposed that it was inherited and that it predisposed carriers to leukaemia. Although this speculation was later disproved, it led to critically important studies on patients with Down's syndrome, who were known to have an inherited predisposition to leukaemia.

In 1964, shortly after Blumberg moved from the NIH to the Fox Chase Cancer Center in Philadelphia, Pennsylvania, he and his colleague Tom London discovered that although the Australia antigen was found at high prevalence in Down's patients as a group, it was not present in newborns. Rather than being inherited, the presence of the antigen in the blood of a Down's patient seemed to correlate with their living in a mental-health

institution. This was the first clue that the Australia antigen might be related to an infectious agent. The specific infection was identified as hepatitis B when two Down's syndrome patients and an investigator in Blumberg's lab were found to carry the antigen only after developing acute hepatitis.

The US company Abbott Laboratories developed a test to screen blood donors for the antigen, which dramatically reduced



the incidence of hepatitis B associated with blood transfusions. The test also identified pregnant women who carried the virus and led to measures that reduced mother-to-infant transmission by 90%.

In the late 1960s, Blumberg and Irving Millman, also at Fox Chase, postulated that the small, non-infectious Australia antigen particle could be separated from the virus and serve as a vaccine. A hepatitis B vaccine — initially produced from the plasma of infected individuals but now manufactured through genetic engineering — was developed by Maurice Hilleman at Merck and has prevented millions of hepatitis B infections. It was also the first cancer vaccine, as hepatitis B virus is a leading cause of liver cancer. For his discoveries, Blumberg received the Nobel Prize in Physiology or Medicine in 1976.

Born in New York City, Blumberg's early schooling was at the Orthodox Yeshiva of Flatbush, which perhaps accounted for his scholarly and deliberative, Talmudic approach to problems in science. After high school, he enlisted in the navy in 1943. A few years later, he pursued a degree in physics at Union College in upstate New

York and then graduate work at Columbia University in New York City. At Columbia, he first studied mathematics but then transferred to become a medical student at the College of Physicians and Surgeons.

It was at medical school that Blumberg's fascination with population genetics and disease took hold. He spent several months in northern Suriname in South America. A richly heterogeneous population had been brought to the country centuries earlier to work in the sugar plantations and Blumberg was struck by the enormous variation in how these people responded to disease. Yet his life took a few more twists and turns before the Australia antigen transformed it. After completing a medical residency at Bellevue Hospital in New York and a clinical fellowship in arthritis at the Columbia-Presbyterian Medical Center, he moved to Britain. He pursued a PhD in the Department of Biochemistry at the University of Oxford on the physical and biochemical characteristics of hyaluronic acid — a key component of connective, epithelial and neural tissues.

Barry's life after he received the Nobel prize was just as varied. He travelled in China, India and Africa to investigate liver cancer and promote the hepatitis B vaccine. In 1989 he was elected Master of Balliol College, Oxford, where, as he put it, he enjoyed "no power, but a great deal of influence". After five years at Balliol, he took a teaching position in medical anthropology at Stanford University in California. There he attended a NASA conference, became captivated by the possibilities of extraterrestrial life, and within months was appointed director of the NASA Astrobiology Institute, a post he held for five years. It was shortly after a return visit to NASA for a keynote lecture that Barry succumbed to a heart attack. He died with his boots on at age 85 on 5 April.

Barry was complex and brilliant, imaginative and adventurous, tenacious and dedicated, a deeply philosophical man of eclectic interests and myriad accomplishments — a Nobel man. ■

Harvey Alter worked with Baruch Blumberg at the US National Institutes of Health and was co-discoverer of the Australia antigen. He is currently Distinguished NIH Investigator in the Department of Transfusion Medicine, Clinical Center, NIH, Bethesda, Maryland, USA.
e-mail: halter@dtm.cc.nih.gov

AP PHOTO/E. ADAMS

Degrees of control

One might expect that social networks would generally be harder to control than naturally occurring systems such as biological networks. But this is not so, according to a new study. [SEE ARTICLE P.167](#)

MAGNUS EGERSTEDT

Networks can be found all around us. Examples include social networks (both online and offline), mobile sensor networks and gene regulatory networks. Such constructs can be represented by nodes and by edges (connections) between the nodes. The nodes are individual decision makers, for instance people on the social-networking website Facebook or DNA segments in a cell. The edges are the means by which information flows and is shared between nodes. But how hard is it to control the behaviour of such complex networks? On page 167 of this issue, Liu *et al.*¹ show that the answer to this question is anything but intuitive.

The flow of information in a network is what enables the nodes to make decisions or to update internal states or beliefs — for example, an individual's political affiliation or the proteins being expressed in a cell. The result is a dynamic network, in which the nodes' states evolve over time. The overall behaviour of such a dynamic network depends on several factors: how the nodes make their decisions and update their states; what information is shared between the edges; and what the network itself looks like — that is, which nodes are connected by edges.

Imagine that you want to start a trend by influencing certain individuals in a social network, or that you want to propagate a drug through a biological system by injecting the drug at particular locations. Two obvious questions are: which nodes should you pick, and how effective are these nodes when it comes to achieving the desired overall behaviour? If the only important factor is the overall spread of information, these questions are related to the question of finding and characterizing effective decision-makers. However, the nodes' dynamics (how information is used for updating the internal states) and the information flow (what information is actually shared) must also be taken into account. In their study, Liu and co-workers¹ do just this by combining the principles of network science with tools found traditionally in the domain of control theory^{2,3}.

Central to the question of how information, injected at certain key locations, can be used to steer the overall system towards some desired performance is the notion of



Figure 1 | Tough job. Liu *et al.*¹ show that complex networks such as biological networks, metaphorically depicted by this locust swarm, are not at all easy to control.

controllability — a measure of what states can be achieved from a given set of initial states. Different dynamical systems have different levels of controllability. For example, a car without a steering wheel cannot reach the same set of states as a car with one, and, as a consequence, is less controllable.

Liu and colleagues¹ found that, for several types of network, controllability is connected to a network's underlying structure^{4–6}. The authors identified what driver nodes — those into which control inputs are injected — can direct the network to a given behaviour. The surprising result is that driver nodes tend to avoid the network hubs. In other words, centrally located nodes are not necessarily the best ones for influencing a network's performance. So for social networks, for example, the most influential members may not be those with the most friends.

The result of this type of analysis^{1,4} is that it is possible to determine how many driver nodes are needed for complete control over a network. Liu *et al.* do this for several real networks, including gene regulatory networks for controlling cellular processes, large-scale data networks such as the World Wide Web, and social networks. We have a certain intuition about how

hard it might be to control such networks. For instance, one would expect cellular processes to be designed to make them amenable to control so that they can respond swiftly to external stimuli, whereas one would expect social networks to be more likely to resist being controlled by a small number of driver nodes.

It turns out that this intuition is entirely wrong. Social networks are much easier to control than biological regulatory networks, in the sense that fewer driver nodes are needed to fully control them — that is, to take the networks from a given configuration to any desired configuration. Liu and colleagues find that, to fully control a gene regulatory network, roughly 80% of the nodes should be driver nodes. By contrast, for some social networks only 20% of the nodes are required to be driver nodes. What's more, the authors show that engineered networks such as power grids and electronic circuits are overall much easier to control than social networks and those involving gene regulation. This is due to both the increased density of the interconnections (edges) and the homogeneous nature of the network structure.

These startling findings¹ significantly further our understanding of the fundamental

AP PHOTO/J. ACOSTA

properties of complex networks. One implication of the study is that both social networks and naturally occurring networks (Fig. 1), such as those involving gene regulation, are surprisingly hard to control. To a certain extent this is reassuring, because it means that such networks are fairly immune to hostile takeovers: a large fraction of the network's nodes must be directly controlled for the whole of it to change. By contrast, engineered networks are generally much easier to control, which may or may not be a good thing, depending on who is trying to control the network. ■

Magnus Egerstedt is in the School of

Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.
e-mail: magnus@gatech.edu

1. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. *Nature* **473**, 167–173 (2011).
2. Mesbahi, M. & Egerstedt, M. *Graph Theoretic Methods in Multiagent Networks* (Princeton Univ. Press, 2010).
3. Kalman, R. E. *J. Soc. Indus. Appl. Math. Ser. A* **1**, 152–192 (1963).
4. Rahmani, A., Ji, M., Mesbahi, M. & Egerstedt, M. *SIAM J. Contr. Optim.* **48**, 162–186 (2009).
5. Tanner, H. G. *43rd IEEE Conf. Decision Contr.* **3**, 2467–2472 (2004).
6. Lin, C.-T. *IEEE Trans. Automat. Contr.* **19**, 201–208 (1974).

CANCER

The flipside of Notch

Mutations that lead to increased activity of the Notch signalling pathway are well defined in human cancer. New work implicates decreased activity of this pathway in a type of blood cancer. SEE LETTER P.230

DEMETRIOS KALAITZIDIS
& SCOTT A. ARMSTRONG

Some of the most common and well studied mutations in human cancers affect signal-transduction pathways. For instance, mutations that lead to increased activity of the receptor protein Notch are frequently found in a type of blood cancer called T-cell acute lymphoblastic leukaemia/lymphoma¹. On page 230 of this issue, Klinakis *et al.*² report that mutations that lead to reduced activity of this protein are associated with another human blood cancer, chronic myelomonocytic leukaemia. This finding suggests that Notch can have either an oncogenic or a tumour-suppressive effect in blood cancers.

The Notch signalling pathway is evolutionarily conserved and has crucial roles in the development and maintenance of embryonic and adult tissues. Notch signalling is initiated when one cell expressing the appropriate ligand interacts with another cell expressing a Notch receptor. Ligand–receptor binding leads to a series of steps involving Notch processing. One such step requires the γ -secretase enzyme complex, which, through protein cleavage, generates a portion of the Notch receptor — called the Notch intracellular domain (NICD) — that is no longer bound to the cell membrane and that relocates to the nucleus (Fig. 1).

In the nucleus, the NICD interacts with DNA-bound protein factors (CSL/CBF1/RBPj γ) and recruits MAML proteins to modulate the expression of many genes³. One of the genes is the Notch target *Hes1*, whose increased

expression is part of the mechanism by which Notch signalling influences cellular physiology. The functions of the Notch pathway are highly cell-type dependent in different embryonic and adult tissues, as well as in cancers⁴. It therefore seems likely that Notch regulates diverse context-specific gene-expression programs that we are just beginning to understand.

To investigate the role of Notch in the haematopoietic system, Klinakis *et al.*²

specifically inactivated the Nicastrin gene in mouse blood cells. (Nicastrin is an essential component of the γ -secretase complex and so is required for the Notch-pathway function.) Surprisingly, the mice died relatively quickly — 20 weeks after birth — from a blood disorder similar to human chronic myelomonocytic leukaemia.

The γ -secretase complex has other functions besides processing Notch⁵. However, the authors confirm the significance of losing Notch signalling by showing that deletion of just the Notch1 and Notch2 receptors from blood cells is sufficient to produce the same cancer in mice. In addition, activation of the Notch pathway in cells lacking Nicastrin ameliorated the leukaemia, further supporting the crucial role of the Notch pathway.

Klinakis and colleagues also show that the effects of Notch loss on blood cells is cell-autonomous — that is, the cancer is due to the loss of Notch function in blood cells and not to its effects on other organs that then feed back to blood cells. This is an important demonstration, because the disruption of Notch signalling in mouse skin also leads to blood disorders in a non-cell-autonomous manner^{6,7}.

Klinakis *et al.* further report that Notch signalling actively represses a gene-expression program in blood stem and progenitor cells that is associated with differentiation of these cells along the myeloid lineage. Thus, loss of Notch signalling seems to 'rewire' early blood cells to inappropriately express genes specifying a myelomonocytic fate that, in mouse models, leads to leukaemia.

In addition to defining a new role for Notch signalling as a suppressor of leukaemia

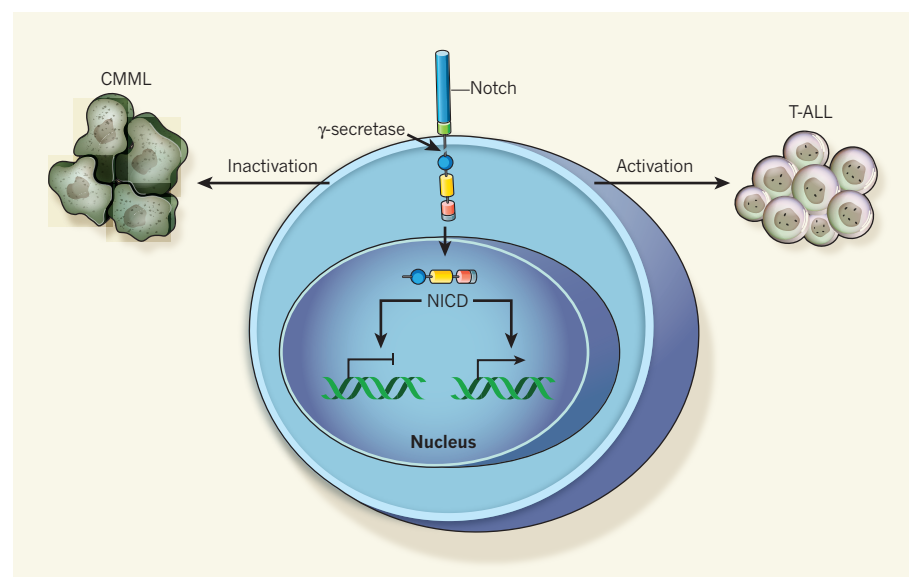


Figure 1 | The Notch signalling pathway and blood disorders. On interaction with an appropriate ligand (not shown) the Notch receptor is processed by the γ -secretase complex to form an intracellular domain (NICD), which accumulates in the nucleus to modulate gene expression. Activating mutations in Notch receptors have been described in T-cell acute lymphoblastic leukaemias/lymphomas (T-ALL), making the receptors an attractive drug target for this cancer. But Klinakis *et al.*² ascribe a tumour-suppressor role for Notch in another blood cancer, chronic myelomonocytic leukaemia (CMML).

development, the authors provide compelling *in vitro* data that these effects are mediated by the Notch transcriptional co-activator MAML1 and by the target gene *Hes1*. But previous studies^{8,9} reported that perturbing Notch transcriptional effectors in mice did not lead to leukaemia during the observation period. Together, these data point to the potential existence of non-canonical Notch effectors in blood stem cells, but more work is needed to clarify these discrepancies.

Klinakis *et al.*² extend their work to human leukaemias harbouring inactivating mutations in members of the Notch pathway. Remarkably, they find six mutations in several genes encoding Notch-pathway members in five out of 42 samples from patients with chronic myelomonocytic leukaemia. In *in vitro* experiments, these mutations disrupted Notch signalling. Furthermore, the authors show that Notch mutations coexist with mutations in pathways previously described in this cancer (those involving the proteins RAS, JAK2 and TET2). This observation indicates cooperativity between Notch loss and perturbations in other signalling pathways implicated in blood cancers, but it needs to be verified in larger collections of patients' samples — work that should be accompanied by functional studies.

Although decreased Notch function has been shown^{10–12} to promote tumour development in specific mouse models, Klinakis and colleagues' work is important because it implicates decreased Notch function in human cancer development. Moreover, this paper² highlights the importance of using mouse models, because work on these animals not only sheds light on gene function in physiological processes, it also helps us to understand human cancer genetics and, potentially, allows the design of new treatments.

This study also underscores the complicated part played by cell-signalling pathways in human disease and the challenges that researchers face in predicting suitable therapeutic targets. Although the Notch pathway is certainly a prime drug target in T-cell acute lymphoblastic leukaemias/lymphomas, the therapeutic window may be dictated by the effects of Notch-pathway inhibition on other cell types, such as myeloid progenitors, as described here. Further investigation of the intricacies of Notch signalling in unaffected and malignant tissues will help to determine the best approaches to manipulating this pathway for optimal therapeutic response. ■

Demetrios Kalaitzidis and Scott A.

Armstrong are at Children's Hospital Boston, and Dana Farber Cancer Institute, Division of Hematology/Oncology, Boston, Massachusetts 02115, USA.

e-mails: demetrios.kalaitzidis@childrens.harvard.edu; scott.armstrong@childrens.harvard.edu

1. Aster, J. C., Blacklow, S. C. & Pear, W. S. *J. Pathol.* **223**, 263–274 (2011).
2. Klinakis, A. *et al. Nature* **473**, 230–233 (2011).
3. Kopan, R. & Ilagan, M. X. G. *Cell* **137**, 216–233 (2009).
4. Koch, U. & Radtke, F. *Curr. Top. Dev. Biol.* **92**, 411–455 (2010).
5. De Strooper, B., Vassar, R. & Golde, T. *Nature Rev. Neurol.* **6**, 99–107 (2010).

6. Demehri, S. *et al. PLoS Biol.* **6**, e123 (2008).
7. Dumortier, A. *et al. PLoS One* **5**, e9258 (2010).
8. Duncan, A. W. *et al. Nature Immunol.* **6**, 314–322 (2005).
9. Maillard, I. *et al. Cell Stem Cell* **2**, 356–366 (2008).
10. Demehri, S., Turkoz, A. & Kopan, R. *Cancer Cell* **16**, 55–66 (2009).
11. Liu, Z. *et al. J. Clin. Invest.* **121**, 800–808 (2011).
12. Hanlon, L. *et al. Cancer Res.* **70**, 4280–4286 (2010).

ASTROPHYSICS

Era of the compact disk

Some of the strangest galaxies in the Universe just got stranger. It seems that many galaxies in the early Universe not only packed a huge number of stars into a very small volume, but were also rotating rapidly.

PIETER VAN DOKKUM

By historical standards the Universe is a boring place. Most galaxies have settled down into the routine of middle age and have all but given up on exciting activities such as forming new stars or fuelling black holes. Studying the light of very distant galaxies with powerful telescopes, astronomers are finding that things were a lot more interesting in the first third of the Universe's history. The most striking aspect of galaxies in those early epochs is their variety^{1,2}. Some were forming stars hundreds of times faster than the Milky Way; others were dominated by accretion flows onto extremely massive black holes in their centres; and still others were crashing into one another to form larger galaxies. At the same time, a class of compact, massive galaxies existed

that had already stopped forming new stars. Writing in *The Astrophysical Journal*, van der Wel *et al.*³ now argue that the majority of these galaxies were disk-shaped.

These compact, massive galaxies are among the most puzzling objects in the young Universe. They seem to have the same number of stars as fully grown galaxies in the present-day Universe. However, their sizes are four to five times smaller and their densities are a hundred times larger^{4,5} than their present-day counterparts. The surprising discovery of these galaxies a few years ago spurred a flurry of studies, which aim to understand how these massive, compact galaxies formed and how they subsequently managed to grow into the much puffier galaxies we have today.

The installation of a new, very sensitive camera on the Hubble Space Telescope in



Figure 1 | Formation of a compact disk galaxy. According to van der Wel *et al.*³, most compact massive galaxies in the young Universe were disk-shaped. During their formation, these galaxies probably resembled the nearby galaxy M82, seen here. The blue light is from young stars in a spinning disk. The red light is from ionized hydrogen gas that is being expelled from the galaxy as a result of supernova explosions. Compact disk galaxies forming in the young Universe may have looked even more spectacular than M82, as they were about the same size but ten times more massive.

2009 offers astronomers the ability to study these remarkable galaxies in much greater detail than before. In their study, van der Wel and colleagues³ went beyond a simple size measurement and determined the morphologies of 14 compact galaxies at a time when the Universe was only about 3 billion years old. The authors concentrated on the degree of flattening of the galaxies. Today's massive galaxies are essentially big balls of stars and tend to be nearly round, whereas lower-mass galaxies such as the Milky Way tend to be elongated because their structure is dominated by a flat, spinning disk of gas and stars. Confirming earlier studies^{5,6} the authors find that a subset of the massive compact galaxies appear highly flattened. Taking into account that some of the apparently round galaxies could be flat disks seen from the top rather than from the side, van der Wel and co-authors argue that most compact galaxies in the young Universe could in fact be spinning disks (Fig. 1).

The case is not yet conclusively proved, because a much larger sample of galaxies needs to be studied to robustly measure the average elongation of compact galaxies. Furthermore, the galaxies may be cigar-shaped (prolate spheroids) rather than disk-shaped. Van der Wel *et al.* argue that prolate galaxies are rare in today's Universe, but so are rapidly rotating, extremely compact disks. The real test is whether these galaxies do, in fact, rotate, which would confirm that they are disks. Such a test is just beyond the capabilities of today's telescopes⁷, but it may be possible to measure the rotation of a similar object that is closer to us, or of a galaxy that happens to be gravitationally lensed — that is, has its light bent and magnified — by a foreground object.

If it is confirmed that these distant galaxies are indeed disks, they would be unlike anything seen in the Universe today. On the basis of the sizes of the galaxies and their masses, the implied rotation speed would be an astonishing 700 kilometres per second. To put this into perspective, the Sun moves around the centre of the Milky Way at a relatively sedate pace of about 230 kilometres per second. This is all the more remarkable given that today's descendants of the compact galaxies are thought to be elliptical galaxies, which do not rotate much at all. It may be that the disks did not survive for long. They may have been destroyed in collisions with other galaxies, which probably occurred frequently during the 10 billion years separating compact galaxies from elliptical galaxies. Some of the disk signatures might be expected to survive this onslaught, and in this context it is interesting that the central regions of elliptical galaxies are often kinematically distinct from the outer regions^{8,9}.

The existence of compact disks of stars in the young Universe would also imply the existence of similarly compact disks of gas at even earlier times — in the first billion years after the Big Bang. Given their densities, these gas

disks would mostly have consisted of molecular hydrogen and helium, which would in turn imply that they converted gas to stars at a ferocious rate while feeding rapidly growing black holes in their centres. It may be possible to detect and characterize these truly spectacular early phases of massive galaxy formation with the Atacama Large Millimeter Array, which is under construction in Chile. ■

Pieter van Dokkum is in the Department of Astronomy, Yale University, New Haven, Connecticut 06511, USA.
e-mail: pieter.vandokkum@yale.edu

CELL SIGNALLING

Why fasting worms age slowly

Lipids of the *N*-acylethanolamine family mediate cell signalling across a wide range of organisms. In nematode worms, they translate food availability into fundamental choices about development that affect lifespan. [SEE LETTER P.226](#)

LUCIANO DE PETROCELLIS
& VINCENZO DI MARZO

Whether prolonged dietary restriction can extend human life remains controversial. But this matter has been settled in the case of some invertebrates, including the widely investigated nematode *Caenorhabditis elegans*¹. A study by Lucanic *et al.*² on page 226 of this issue now suggests that decreased signalling by small lipids called *N*-acylethanolamines is responsible for the lifespan-extending effects of dietary restriction in *C. elegans*.

N-Acylethanolamines (NAEs) were first identified in the 1950s, but evidence for their signalling role emerged only when a member of this lipid family, *N*-arachidonylethanolamine (anandamide), was detected in pig brains³ and recognized as an endocannabinoid — an endogenous activator of CB₁ cannabinoid receptors. It is now known that these small lipids, which have fatty-acid chains of varying length and saturation level, are widespread in both the plant and animal kingdoms, and are even found in single-celled protists. In mammals, signalling through CB₁ receptors by anandamide — and another endocannabinoid, 2-arachidonylethanolamine — is essential for stimulating food intake following a short period of food deprivation⁴. This observation provided the rationale for the development of CB₁ blockers as anti-obesity agents⁴.

A relationship between an organism's energy status and its brain levels of endocannabinoids is not restricted to mammals, but is also found

1. Papovich, C. *et al. Astrophys. J.* **640**, 92–113 (2006).
2. Kriek, M., van Dokkum, P. G., Franx, M., Illingworth, G. D. & Magee, D. K. *Astrophys. J.* **705**, L71–L75 (2009).
3. van der Wel, J. *et al. Astrophys. J.* **730**, 38–41 (2011).
4. Daddi, E. *et al. Astrophys. J.* **626**, 680–697 (2005).
5. van Dokkum, P. G. *et al. Astrophys. J.* **677**, L5–L8 (2008).
6. McGrath, E. J., Stockton, A., Canalizo, G., Iye, M. & Maihara, T. *Astrophys. J.* **682**, 303–318 (2008).
7. van Dokkum, P. G., Kriek, M. & Franx, M. *Nature* **460**, 717–719 (2009).
8. Franx, M. & Illingworth, G. D. *Astrophys. J.* **327**, L55–L59 (1988).
9. Krajinović, D. *et al. Mon. Not. R. Astron. Soc.* **390**, 93–117 (2008).

in non-mammalian vertebrates such as fish, in which these compounds also stimulate food intake (Fig. 1). In the primitive invertebrate *Hydra vulgaris*, which does not seem to have a protein related to cannabinoid receptors, anandamide nonetheless inhibits feeding behaviour (mouth opening). And the chordate *Ciona intestinalis*, which represents the phylogenetic branching point between invertebrates and vertebrates, expresses a receptor that is homologous to the human CB₁ receptor. In this animal, synthetic CB₁ agonists retard mouth reopening after closure⁴.

The discovery of anandamide³ ignited interest in other NAEs, although it was immediately clear that most of these compounds do not share anandamide's high affinity for CB₁ receptors. In rodents, *N*-oleoylethanolamine activates PPAR- α , a nuclear receptor that acts as a transcription factor, inhibiting food intake⁵. But apart from studies on this compound, and on the potent anti-inflammatory mediator *N*-palmitoylethanolamine, research in animals on NAEs that do not bind to cannabinoid receptors has been sparse, unlike work on the signalling function of these lipids in plants^{6,7}. Notably, NAE biosynthetic pathways, and the enzymes that inactivate these compounds, are similar in animals and plants⁷.

Lucanic and colleagues² identify NAEs in *C. elegans*, and demonstrate that the abundance of these lipids is reduced under dietary restriction. Moreover, they show that NAE deficiency — due either to impaired biosynthesis or enhanced degradation in mutant worm strains — is sufficient to enhance longevity.



50 Years Ago

The [World Health Organization] programme for the eradication of malaria from the world is reported as having made significant progress in various fields, although handicapped by a persistent shortage of funds. Sixty-one countries are now fully engaged in this programme; in another nineteen, planning has reached an advanced stage, and, in seventeen other areas—mainly in Africa—the Organization is assisting with preliminary work... The World Health Organization's assistance to Governments for the control of yaws, smallpox and tuberculosis, leprosy and various diseases increased significantly during the year. Medical research was intensified on practically all communicable diseases, the main co-ordinated effort being centred on virus diseases and on bilharziasis [schistosomiasis].

From *Nature* 13 May 1961

100 Years Ago

Whatever may be thought of Mr. Willett's so-called daylight-saving scheme, it is impossible not to admire the persistence with which he pursues the idea, and secures support for it from city corporations, town councils, chambers of commerce, members of Parliament, and other people who are attracted by the advantages offered, and do not realise how unscientific the scheme is, or the gravity of the objections to the adoption of a variable standard of time-reckoning. We do not believe for an instant that the Government is likely to give facilities for legislation on the lines of the Summer Season Time Bill... The scheme is unworthy of the dignity of a great nation... We cannot think that the Government will lend its support to proposals which... would make us the laughing-stock of the enlightened people of the world.

From *Nature* 11 May 1911

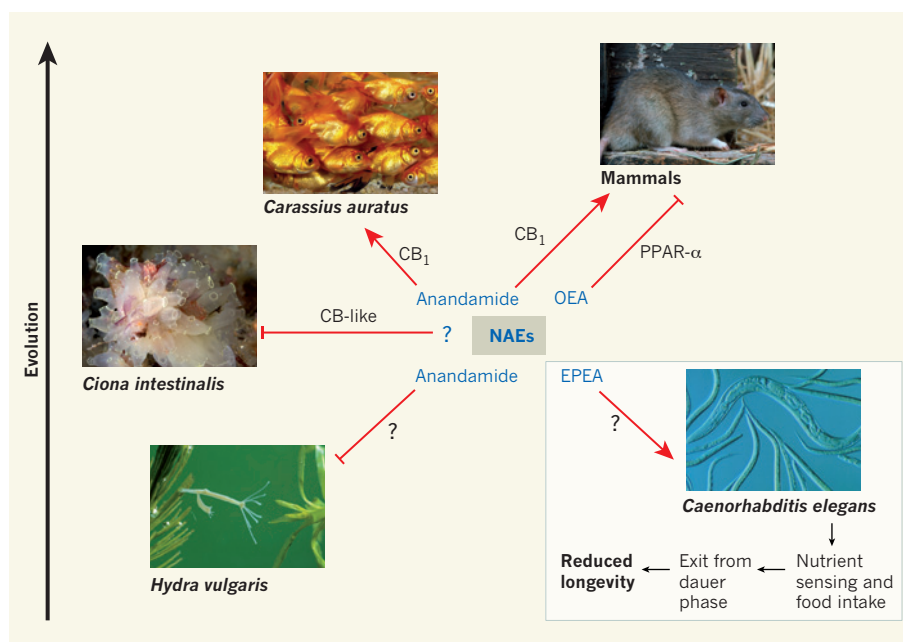


Figure 1 | *N*-Acylethanolamines, feeding behaviour and lifespan. *N*-Acylethanolamines (NAEs) are evolutionarily conserved lipids: they occur in a wide range of organisms, including the coelenterate *Hydra vulgaris*, the chordate *Ciona intestinalis* and the fish *Carassius auratus*. In all cases, these small lipids either inhibit (blunt-ended arrows) or enhance (pointed arrows) feeding behaviours. In mammals, CB₁ and PPAR- α receptors mediate the effects of NAEs such as anandamide and *N*-oleoylethanolamine (OEA), respectively, although NAE receptors in some of the other species are unknown (black question marks). Inset: Lucanic *et al.*² show that in the nematode *Caenorhabditis elegans*, *N*-eicosapentaenoyl ethanolamine (EPEA), the most abundant NAE in this organism, enhances feeding behaviour. This leads to exit of the animals from the dauer phase — which they originally enter when faced with food shortage — and to reduced longevity.

Using other types of mutants, the authors further show that NAE deficiency and dietary restriction extend *C. elegans* lifespan through the same developmental effects (see below and Fig. 1).

Conversely, supplementing the worms' diet with *N*-eicosapentaenoyl ethanolamine (EPEA), the most abundant NAE in *C. elegans*, inhibited dietary-restriction-induced lifespan extension in wild-type worms; it also had the same effect in more long-lived mutants in which nutrient sensing was impaired because they lacked the signalling molecule TOR, which is necessary for food intake. Lucanic *et al.* propose that NAE-mediated signalling, which is presumably acting in the pharynx, coordinates nutrient sensing and energy status with the metabolic and developmental changes that ultimately determine lifespan in worms.

Rodents cope with prolonged semi-starvation by — among other things — reducing the levels of appetite-inducing endocannabinoids in their hypothalamus (the brain area that mediates homeostatic control of energy intake)⁴. Instead, in the face of dietary restriction, *C. elegans* enters a quiescent and low-energy-consuming state known as the dauer phase, in which NAE levels are at their lowest². Lucanic and co-workers² show that administration of exogenous EPEA can bring the nematode out of the dauer phase, and they propose that it does so by providing “a false signal

of high nutrient availability” and by reversing “the metabolic adaptation to reduced food availability that confers lifespan extension”.

Perhaps the only missing information in the researchers' extensive study is the receptor through which EPEA, and possibly other NAEs, produce their effects in *C. elegans* — like most other invertebrates, this animal does not express the equivalent of cannabinoid receptors⁵. The authors did look at ‘indigenous’ members of the several classes of receptor that are activated by these lipids in mammals, including orphan G-protein-coupled receptors, nuclear receptors and ligand-activated ion channels⁹. But genetic deletion of proteins related to such potential NAE targets in *C. elegans* did not affect the ability of EPEA to rescue the worm from the dauer phase², leaving the identity of the receptor mediating NAE signalling in this animal a mystery.

Nevertheless, apart from the discovery of novel molecular mechanisms underlying dietary-restriction-induced developmental changes in *C. elegans*, the elegant work of Lucanic and colleagues' sends another noteworthy general message: across nearly all types of living organism, NAEs are of fundamental importance as signalling molecules that act through different receptors (with sometimes more than one receptor mediating the effects of a particular NAE)⁹, and have diverse biological actions.

FROM BOTTOM: K. WOTHE/MINDEN PICTURES/FLPA; S. STAMMERS/SPL; K. TELNES/IMAGEQUEST/MARINE.COM; G. LACZ/FLPA; REINHARD/ARCO/NATURE/PLCO

As previously suggested¹⁰ — and as supported by the identification of enzymes mediating NAE biosynthesis and degradation in even elementary forms of life — NAEs are phylogenetically ancient. So it could be that, during evolution, when cannabinoid receptors and PPAR- α were first used as signal transducers, they found endogenous NAEs and used them as activators. Regardless of the identity of their receptors, however, the lipids' key role in the control of food intake and development seems to have been conserved from simple organisms to humans (Fig. 1). Thus, finding the molecular targets of NAEs in *C. elegans*

should be made a priority, because these proteins might also have been retained in mammals, with potential implications for medical research in ageing, obesity and associated metabolic disorders. ■

Luciano De Petrocellis and Vincenzo Di Marzo are in the Endocannabinoid Research Group, Institute of Biomolecular Chemistry and Institute of Cybernetics, Consiglio Nazionale delle Ricerche, Via Campi Flegrei 34, 80078 Pozzuoli (NA), Italy.
e-mail: vincenzo.dimarzo@icb.cnr.it

1. Fontana, L., Partridge, L. & Longo, V. D. *Science* **328**, 321–326 (2010).
2. Lucanic, M. et al. *Nature* **473**, 226–229 (2011).
3. Devane, W. A. et al. *Science* **258**, 1946–1949 (1992).
4. Matias, I., Bisogno, T. & Di Marzo, V. *Int. J. Obes.* **30**, S7–S12 (2006).
5. Fu, J. et al. *Nature* **425**, 90–93 (2003).
6. Berdyshev, E. V. et al. *World Rev. Nutr. Diet.* **88**, 207–214 (2001).
7. Kilaru, A. et al. *Chem. Biodivers.* **4**, 1933–1955 (2007).
8. Elphick, M. R. & Egertová, M. *Handb. Exp. Pharmacol.* **168**, 283–297 (2005).
9. Pertwee, R. G. et al. *Pharmacol. Rev.* **62**, 588–631 (2010).
10. McPartland, J. M., Norris, R. W. & Kilpatrick, C. W. *Gene* **397**, 126–135 (2007).

ARCHAEOLOGY

The cost of cultivation

There were probably many reasons for the adoption of agriculture by prehistoric human societies. A fresh perspective comes from a quantitative estimate of the relative productivity of farming and foraging.

GRAEME BARKER

The origins of agriculture are among the 'big questions' about the human past that archaeologists have grappled with throughout the history of the discipline. A study by Samuel Bowles, published in *Proceedings of the National Academy of Sciences*¹, shows that cultivating cereals was not just harder work than hunting and gathering, but was probably less productive. So why did foragers become farmers?

For decades, most archaeologists have agreed that the first farming — the Neolithic Revolution in Gordon Childe's famous phrase² — began soon after the transition from the Pleistocene to the Holocene climatic eras some 12,000 years ago, in a few 'hearths of domestication' such as southwest Asia, China and parts of the Americas; and that over the next few thousand years a variety of agricultural systems replaced foraging (hunting and gathering) in almost every corner of the globe. As revolutions go it may have taken several thousand years, and developed piecemeal across the globe, but it remains one of the most extraordinary transformations in human history³.

So why did most foragers decide to become farmers? Childe emphasized the advantages of farming over foraging: it gave people a reliable food supply, allowing them to settle down, and this, in combination with the possibilities farming created for producing surplus food, provided the springboard for global population growth and

transformations in social complexity that led within a few millennia to urbanism.

Ethnographic studies in the 1960s dramatically changed perceptions of farming's advantages over foraging: even in a marginal environment such as the Kalahari desert, foragers had a secure food base, and to exploit it worked fewer hours than most farmers need to do in non-industrial societies⁴. In recent decades, therefore, archaeologists have theorized about the processes that might have persuaded foragers to become farmers, even though farming probably meant more work, a

less varied diet and more disease, to say nothing of the social stresses of moving from group food-sharing to household-based production. They have variously proposed triggers such as climate change, population pressure and social competition, or combinations of all three, but they have at least agreed that in most situations farming would have been more productive than foraging.

Bowles¹ has added further sauce to the rich mix of theories. His analysis of ethnographic and historic figures for foragers and farmers in traditional societies (that is, using hand tools) suggests that the average productivity from cultivating seed plants (he doesn't deal with tubers) is about three-fifths of the return from foraging wild species. In response to the question 'why did the first farmers farm?', he proposes that many foragers might well have been attracted to undertake some small-scale farming because it just meant forgoing those wild plants or animals with the lowest calorific return rate. Indeed, there are many archaeological examples of foragers who engaged in cultivating plants and/or herding animals for many centuries before developing a significant commitment to farming.

But why develop such a commitment if the returns from farming were likely to be lower than those from foraging? Bowles concludes that foragers might have been drawn into farming because becoming more sedentary would have lowered the costs of child rearing and inevitably stimulated population growth, and because producing surplus food meant increased possibilities for social competition. Both are arguments that would be familiar to Childe, but they still remain more descriptions of what might have happened than explanations of why people took the decisions they did.

Most theorizing about the origins of agriculture has started from the premise that contemporary or ethnographic examples of modern 'traditional' foraging and farming societies are likely guides to how late Pleistocene and early Holocene



Figure 1 | Planting rice in Borneo. Today rice is a staple crop on which millions of people depend. But for many prehistoric societies its role as a high-status, special food may well have been more important than its dietary contribution.

H. BARTON

societies may have functioned in the past, but archaeology increasingly reveals a complicated and ambiguous past — another country in which people did things differently. Late Pleistocene foragers in the islands of southeast Asia were burning vegetation to enhance plant productivity 50,000 years ago^{5,6} and Upper Palaeolithic reindeer hunters had domesticated dogs 30,000 years ago⁷. People in the Near East and China were cultivating wild cereals for thousands of years before the plants acquired the morphologies of modern domesticated crops^{8,9}. In North Africa, the wild Barbary sheep was herded 500 years before people acquired domestic sheep and goats¹⁰. There are other examples of experiments in managing animals and plants that failed to survive into modern farming.

In many 'early farming' societies, the domesticates, and the material culture associated with them (such as sickles in the case of the first cereals in the Near East), were embedded in elaborate social and ritual behaviours, suggesting that motivations for acquiring and using new food resources were highly variable and rarely simply a matter of dietary stress or opportunity^{11,12}. In Borneo, for example, rice was used on a small scale for thousands of years by people who relied on hunting and the management of palms and tubers; difficult to grow and high risk, rice only became a food staple there a few centuries ago (Fig. 1). But before that, it almost certainly had a critical social role as a high-status special food with magical properties¹³.

It is probably illusory to expect some kind of systematic global causality for the first farming¹⁴ — people adopted new technologies and foods in many different ways, at many different timescales, and surely for widely differing reasons. Bowles's paper¹ is further evidence that we should not impose Enlightenment notions of rationality on decision-making by prehistoric foragers. Such decisions had many unintended consequences, often ones that in time changed the social landscape irrevocably and made once-optional additions to forager lifestyles obligatory components of new ways of living. ■

Graeme Barker is in the McDonald Institute for Archaeological Research and the Department of Archaeology, University of Cambridge, Downing Street, Cambridge CB2 3ER, UK.
e-mail: gb314@cam.ac.uk

1. Bowles, S. *Proc. Natl Acad. Sci. USA* **108**, 4760–4765 (2011).
2. Childe, V. G. *Man Makes Himself* (Watts, 1936).
3. Barker, G. *The Agricultural Revolution in Prehistory: Why Did Foragers Become Farmers?* (Oxford Univ. Press, 2006).
4. Lee, R. B. & DeVore, I. (eds) *Man the Hunter* (Aldine, 1968).
5. Barker, G. *et al. J. Hum. Evol.* **52**, 243–261 (2007).
6. Summerhayes, G. R. *et al. Science* **330**, 78–81 (2010).

7. Germonpré, M. *et al. J. Archaeol. Sci.* **36**, 473–490 (2009).
8. Fuller, D. Q. *Ann. Bot.* **100**, 903–924 (2007).
9. Jones, M. K. & Liu, X. *Science* **324**, 730–731 (2009).
10. di Lernia, S. in *Before Food Production in North Africa* (eds di Lernia, S. & Manzi, G.) 113–126 (ABACO, 1998).
11. Barton, H. *Curr. Anthropol.* **50**, 673–675 (2009).

12. Finlayson, B. & Warren, G. (eds) *Landscapes in Transition* (Oxbow, 2010).
13. Barker, G., Hunt, C. & Carlos, J. in *Why Cultivate? Anthropological and Archaeological Approaches to Foraging–Farming Transitions in Southeast Asia* (eds Barker, G. & Janowski, M.) 59–72 (McDonald Inst. Archaeol. Res., 2011).
14. Fuller, D., Allaby, R. & Stevens, C. *World Archaeol.* **42**, 13–28 (2010).

QUANTUM PHYSICS

Keep your feet on the ground

Some complex problems in physics can be recast as finding the ground state of an interacting quantum system. Not getting excited along the way can be the challenging part. [SEE LETTER P.194](#)

WILLIAM D. OLIVER

Information processing based on quantum-mechanical interactions has the potential to yield a new and powerful set of computational tools and capabilities. In essence, replacing the classical bits of information in today's electronic devices with quantum bits (qubits) enables a computation to proceed on the basis of the rules of quantum mechanics. As one might surmise, however, if the quantum nature of this computational evolution were somehow corrupted along the way, the entire computation would be brought into question. On page 194 of this issue, Johnson *et al.*¹ demonstrate that the evolution of a superconducting system of eight qubits is indeed consistent with the principles of quantum mechanics when a particular protocol called quantum annealing is implemented.

The solution for certain, challenging, optimization problems can be determined from the ground (minimal-energy) state of a system of interacting spins. Finding this ground state is itself a computationally demanding problem, and a form of adiabatic quantum computation called quantum annealing is one proposed means to do it^{2,3}. Conceptually, an adiabatic process works as follows. The system starts in a known ground state with the interactions between neighbouring spins effectively turned off. The interactions are then slowly turned on such that the system evolves adiabatically — that is, without ever leaving its instantaneous ground state during this evolution². If the system is never excited, always remaining in the instantaneous ground state, then it will surely end in the final, desired interacting ground state and we have the answer.

To illustrate how this process unfolds, consider a corrugated egg carton, with the ridges and furrows creating potential wells, and an egg placed in one of these wells.

As the interactions are turned on, the egg carton slowly changes shape (never so fast that it jostles the egg into an adjacent well), and one of the wells becomes deeper than the others; this deeper well is the true ground state. A quantum egg will keep its cool and quantum mechanically tunnel through the ridges to reach this global minimum, whereas a classical thermal egg must resort to jumping excitedly over them. In the language of quantum annealing, the system continuously seeks the lowest-energy configuration by means of quantum tunnelling mediated by quantum fluctuations. This process should be distinguished from classical thermal annealing, in which thermal fluctuations lead the way.

In their study, Johnson *et al.*¹ employ a one-dimensional Ising spin system, a chain of eight spins which can interact with one another in a nearest-neighbour pairwise manner. The spins are realized with superconducting flux qubits⁴, manufactured artificial spins whose values 'spin-up' and 'spin-down' correspond to clockwise and counter-clockwise circulating superconducting currents, which are respectively associated with the left and right wells of the qubit's potential energy diagram (Fig. 1). The authors went to great lengths to make the various qubit parameters tuneable, so that they could carefully calibrate the settings for each individual spin, as well as tune the spin–spin coupling strength and its sign (negative for ferromagnetic coupling and positive for anti-ferromagnetic coupling).

To demonstrate the distinction between quantum and thermal annealing, Johnson *et al.* leveraged ideas from the domain of macroscopic quantum tunnelling^{5,6}. Consider first a single qubit and its double-well potential. The qubit starts in a symmetrical configuration with a low-potential tunnel barrier and an equal population in both wells — that is, with equal probability for spin-up or spin-down

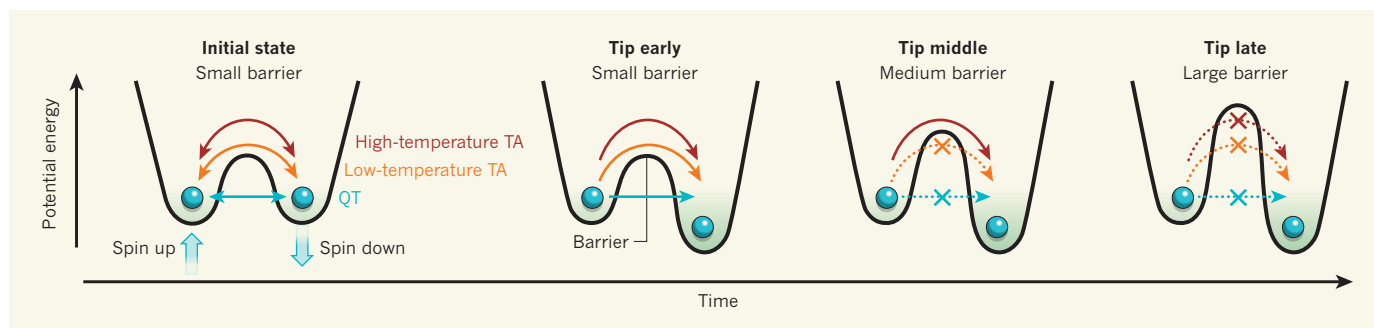


Figure 1 | Finding the ground state for a single-qubit system by quantum annealing¹. A single qubit potential is represented by two wells separated by a barrier. The wells correspond to the qubit's spin-up and spin-down states. Initially, the potential is symmetrical and has a low barrier, allowing both thermal activation (TA) over the barrier and quantum tunnelling (QT) through the barrier. This distributes population equally in each well (the probability for spin-up and spin-down states to occur is the same). During the quantum-annealing protocol, the barrier height is slowly increased. At

various times and temperatures, the double well is 'tipped' to see whether the higher-energy left-well population can find its way to the lower-energy right well. When tipped early, both QT and TA are allowed and the redistribution of population occurs at all temperatures. For medium barriers, the population cannot redistribute ('freezes') for low temperatures, but is thermally activated for high temperatures. For large barriers, no redistribution can occur. By finding the population 'freeze time' for various times and temperatures, it is possible to distinguish between QT and TA.

states to occur. As the tunnel barrier is continuously raised, the rates for quantum tunnelling (through the barrier) and thermal activation (over the barrier) continuously decrease until they effectively stop, freezing the population in place. The freezing for quantum tunnelling should depend primarily on the barrier, whereas the freezing for thermal activation should also scale with temperature (which provides the energy to 'jump over' the barrier). Which of the two processes causes the population to freeze in this system?

To answer this question, the authors raise the potential energy of the left well with respect to the right well at several times during the system's evolution to test whether, at these various times, the left-well population can find its way through the ever-increasing barrier to the lower-energy right well. They do this exercise not only as a function of time but also as a function of temperature (Fig. 1).

They find that the population freeze time (the time at which the barrier becomes so large that subsequently raising the left well has no further effect on the population distribution) saturates to a constant value as the temperature is reduced below 45 millikelvin, consistent with a quantum-tunnelling picture. Had a classical thermal activation picture been correct, one would expect the population to change continuously with temperature according to the value of the temperature and the height of the barrier. This conclusion presumes that the temperature itself is not stuck at 45 millikelvin. As a check, Johnson *et al.* show that at very early times, while the barrier is still low enough to allow thermal activation to occur, the freeze time does not saturate but instead scales with temperature down to 20 millikelvin. This serves as evidence that the population-freezing effect is not due to an inability to cool the device.

Johnson and colleagues¹ then proceed to apply the quantum-annealing technique to the eight-qubit system. In this case, the first and

last spins in the chain are forced to be spin-up and spin-down, respectively, while the remaining six spins are allowed to take on either value. With the spin-spin interactions turned on, the system is 'frustrated': although all spins would prefer to align with their neighbours (the authors use ferromagnetic coupling), this is not possible given that the first and last spins are fixed in opposite directions. As a result, a domain wall (an adjacent spin-up and spin-down) must appear somewhere, and its initial placement along the chain is equally likely. As before, the authors proceed to increase the barriers for each spin and, again, at various times and temperatures, raise the left wells for the inner six spins. Doing so makes one particular placement of the domain wall assume a lower energy than the others and, as before, the test is to see if the initial spin arrangements can find their way to this lowest-energy configuration. By finding the freeze time for the eight-qubit spin distribution as a function of temperature, the authors similarly determine that the observed spin-reconfiguration dynamics is most consistent with a quantum-tunnelling picture.

Given the number of dynamical parameters (changing the barrier height, raising the left wells, the time at which the wells are raised, and the temperature), the authors required simulations to draw their conclusions. Both the single-qubit and eight-qubit experiments were inconsistent with a classical model of thermal activation over the barrier to reach the ground state. Rather, the evolution was consistent with a theoretical quantum model that comprised four quantized energy levels in the presence of a thermal bath and that was simulated over a rather incredible 2 million computer-processing hours.

To be clear, this system was not used to perform any computational algorithm. And, whereas computing algorithms based on adiabatic quantum annealing have been demonstrated with small-scale model problems

in nuclear magnetic resonance systems⁷, the practical benefits and scalability of quantum-annealing algorithms for large-scale problems remain unclear^{8,9}. Furthermore, although Johnson and colleagues¹ ruled out classical thermal activation over the barrier, they could not rule out thermal excitation within the potential wells. Such intra-well excitations certainly existed during the procedure employed by the authors (their purpose here was solely to distinguish quantum tunnelling from classical thermal activation over the barrier to reach the ground state), and these additional excitations would certainly complicate (if not prohibit outright) the implementation of an adiabatic quantum-annealing algorithm and reduce its effectiveness.

Nonetheless, the demonstration that an eight-qubit system evolves according to a quantum model in the presence of a thermal bath stands as a technical achievement, and its practical significance will ultimately be judged by the degree to which it enables the authors to address these and related questions. Whatever the outcome may be, there is still much to learn from a system that aims to remain firmly planted on the ground. ■

William D. Oliver is in the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts 02420-9108, USA.
e-mail: oliver@ll.mit.edu

1. Johnson, M. W. *et al.* *Nature* **473**, 194–198 (2011).
2. Farhi, E. *et al.* *Science* **292**, 472–475 (2001).
3. Santoro, G. E., Martoňák, R., Tosatti, E. & Car, R. *Science* **295**, 2427–2430 (2002).
4. Clarke, J. & Wilhelm, F. K. *Nature* **453**, 1031–1042 (2008).
5. Voss, R. F. & Webb, R. A. *Phys. Rev. Lett.* **47**, 265–268 (1981).
6. Devoret, M. H., Martinis, J. M. & Clarke, J. *Phys. Rev. Lett.* **55**, 1908–1911 (1985).
7. Chen, H. *et al.* *Phys. Rev. A* **83**, 032314 (2011).
8. Farhi, E., Goldstone, J. & Gutmann, S. Preprint at <http://arxiv.org/abs/quant-ph/0201031> (2002).
9. Matsuda, Y., Nishimori, H. & Katzgraber, H. G. *N. J. Phys.* **11**, 073021 (2009).

Controllability of complex networks

Yang-Yu Liu^{1,2}, Jean-Jacques Slotine^{3,4} & Albert-László Barabási^{1,2,5}

The ultimate proof of our understanding of natural or technological systems is reflected in our ability to control them. Although control theory offers mathematical tools for steering engineered and natural systems towards a desired state, a framework to control complex self-organized systems is lacking. Here we develop analytical tools to study the controllability of an arbitrary complex directed network, identifying the set of driver nodes with time-dependent control that can guide the system's entire dynamics. We apply these tools to several real networks, finding that the number of driver nodes is determined mainly by the network's degree distribution. We show that sparse inhomogeneous networks, which emerge in many real complex systems, are the most difficult to control, but that dense and homogeneous networks can be controlled using a few driver nodes. Counterintuitively, we find that in both model and real systems the driver nodes tend to avoid the high-degree nodes.

According to control theory, a dynamical system is controllable if, with a suitable choice of inputs, it can be driven from any initial state to any desired final state within finite time^{1–3}. This definition agrees with our intuitive notion of control, capturing an ability to guide a system's behaviour towards a desired state through the appropriate manipulation of a few input variables, like a driver prompting a car to move with the desired speed and in the desired direction by manipulating the pedals and the steering wheel. Although control theory is a mathematically highly developed branch of engineering with applications to electric circuits, manufacturing processes, communication systems^{4–6}, aircraft, spacecraft and robots^{2,3}, fundamental questions pertaining to the controllability of complex systems emerging in nature and engineering have resisted advances. The difficulty is rooted in the fact that two independent factors contribute to controllability, each with its own layer of unknown: (1) the system's architecture, represented by the network encapsulating which components interact with each other; and (2) the dynamical rules that capture the time-dependent interactions between the components. Thus, progress has been possible only in systems where both layers are well mapped, such as the control of synchronized networks^{7–10}, small biological circuits¹¹ and rate control for communication networks^{4–6}. Recent advances towards quantifying the topological characteristics of complex networks^{12–16} have shed light on factor (1), prompting us to wonder whether some networks are easier to control than others and how network topology affects a system's controllability. Despite some pioneering conceptual work^{17–23} (Supplementary Information, section II), we continue to lack general answers to these questions for large weighted and directed networks, which most commonly emerge in complex systems.

Network controllability

Most real systems are driven by nonlinear processes, but the controllability of nonlinear systems is in many aspects structurally similar to that of linear systems³, prompting us to start our study using the canonical linear, time-invariant dynamics

$$\frac{dx(t)}{dt} = Ax(t) + Bu(t) \quad (1)$$

where the vector $\mathbf{x}(t) = (x_1(t), \dots, x_N(t))^T$ captures the state of a system of N nodes at time t . For example, $x_i(t)$ can denote the amount

of traffic that passes through a node i in a communication network²⁴ or transcription factor concentration in a gene regulatory network²⁵. The $N \times N$ matrix A describes the system's wiring diagram and the interaction strength between the components, for example the traffic on individual communication links or the strength of a regulatory interaction. Finally, B is the $N \times M$ input matrix ($M \leq N$) that identifies the nodes controlled by an outside controller. The system is controlled using the time-dependent input vector $\mathbf{u}(t) = (u_1(t), \dots, u_M(t))^T$ imposed by the controller (Fig. 1a), where in general the same signal $u_i(t)$ can drive multiple nodes. If we wish to control a system, we first need to identify the set of nodes that, if driven by different signals, can offer full control over the network. We will call these 'driver nodes'. We are particularly interested in identifying the minimum number of driver nodes, denoted by N_D , whose control is sufficient to fully control the system's dynamics.

The system described by equation (1) is said to be controllable if it can be driven from any initial state to any desired final state in finite time, which is possible if and only if the $N \times NM$ controllability matrix

$$C = (B, AB, A^2B, \dots, A^{N-1}B) \quad (2)$$

has full rank, that is

$$\text{rank}(C) = N \quad (3)$$

This represents the mathematical condition for controllability, and is called Kalman's controllability rank condition^{1,2} (Fig. 1a). In practical terms, controllability can be also posed as follows. Identify the minimum number of driver nodes such that equation (3) is satisfied. For example, equation (3) predicts that controlling node x_1 in Fig. 1b with the input signal u_1 offers full control over the system, as the states of nodes x_1, x_2, x_3 and x_4 are uniquely determined by the signal $u_1(t)$ (Fig. 1c). In contrast, controlling the top node in Fig. 1e is not sufficient for full control, as the difference $a_{31}x_2(t) - a_{21}x_3(t)$ (where a_{ij} are the elements of A) is not uniquely determined by $u_1(t)$ (see Fig. 1f and Supplementary Information section III.A). To gain full control, we must simultaneously control node x_1 and any two nodes among $\{x_2, x_3, x_4\}$ (see Fig. 1h, i for a more complex example).

To apply equations (2) and (3) to an arbitrary network, we need to know the weight of each link (that is, the a_{ij}), which for most real

¹Center for Complex Network Research and Departments of Physics, Computer Science and Biology, Northeastern University, Boston, Massachusetts 02115, USA. ²Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ³Nonlinear Systems Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁴Department of Mechanical Engineering and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁵Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.

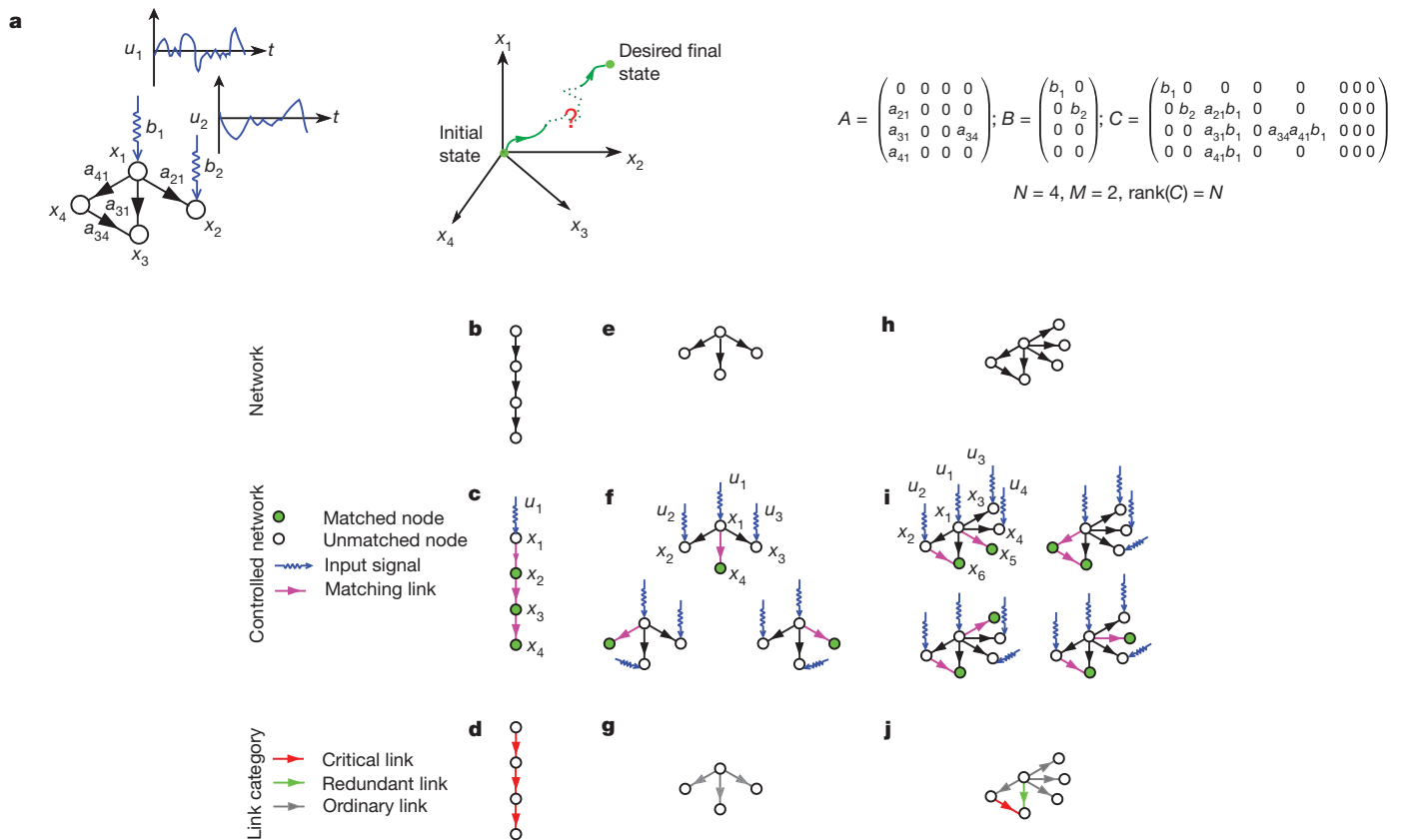


Figure 1 | Controlling a simple network. **a**, The small network can be controlled by an input vector $\mathbf{u} = (u_1(t), u_2(t))^T$ (left), allowing us to move it from its initial state to some desired final state in the state space (right). Equation (2) provides the controllability matrix (C), which in this case has full rank, indicating that the system is controllable. **b**, Simple model network: a directed path. **c**, Maximum matching of the directed path. Matching edges are shown in purple, matched nodes are green and unmatched nodes are white. The unique maximum matching includes all links, as none of them share a common starting or ending node. Only the top node is unmatched, so controlling it yields full control of the directed path ($N_D = 1$). **d**, In the directed path shown in **b**, all links are critical, that is, their removal eliminates our ability to control the network. **e**, Small model network: the directed star. **f**, Maximum matchings of

networks are either unknown (for example regulatory networks) or are known only approximately and are time dependent (for example Internet traffic). Even if all weights are known, a brute-force search requires us to compute the rank of C for $2^N - 1$ distinct combinations, which is a computationally prohibitive task for large networks. To bypass the need to measure the link weights, we note that the system (A, B) is ‘structurally controllable’²⁶ if it is possible to choose the non-zero weights in A and B such that the system satisfies equation (3). A structurally controllable system can be shown to be controllable for almost all weight combinations, except for some pathological cases with zero measure that occur when the system parameters satisfy certain accidental constraints^{26,27}. Thus, structural controllability helps us to overcome our inherently incomplete knowledge of the link weights in A . Furthermore, because structural controllability implies controllability of a continuum of linearized systems²⁸, our results can also provide a sufficient condition for controllability for most non-linear systems³ (Supplementary Information, section III.A).

To avoid the brute-force search for driver nodes, we proved that the minimum number of inputs or driver nodes needed to maintain full control of the network is determined by the ‘maximum matching’ in the network, that is, the maximum set of links that do not share start or end nodes (Fig. 1c, f, i). A node is said to be matched if a link in the maximum matching points at it; otherwise it is unmatched. As we show in the Supplementary Information, the structural controllability

the directed star. Only one link can be part of the maximum matching, which yields three unmatched nodes ($N_D = 3$). The three different maximum matchings indicate that three distinct node configurations can exert full control. **g**, In a directed star, all links are ordinary, that is, their removal can eliminate some control configurations but the network could be controlled in their absence with the same number of driver nodes N_D . **h**, Small example network. **i**, Only two links can be part of a maximum matching for the network in **h**, yielding four unmatched nodes ($N_D = 4$). There are all together four different maximum matchings for this network. **j**, The network has one critical link, one redundant link (which can be removed without affecting any control configuration) and four ordinary links.

problem maps into an equivalent geometrical problem on a network: we can gain full control over a directed network if and only if we directly control each unmatched node and there are directed paths from the input signals to all matched nodes²⁹. The possibility of determining N_D , using this mapping, is our first main result. As the maximum matching in directed networks can be identified numerically in at most $O(N^{1/2}L)$ steps³⁰, where L denotes the number of links, the mapping offers an efficient method to determine the driver nodes for an arbitrary directed network.

Controllability of real networks

We used the tools developed above to explore the controllability of several real networks. The networks were chosen for their diversity: for example, the purpose of the gene regulatory network is to control the dynamics of cellular processes, so it is expected to evolve towards a structure that is efficient from a control perspective, potentially implying a small number of driver nodes (that is, small $n_D \equiv N_D/N$). In contrast, for the World Wide Web or citation networks controllability has no known role, making it difficult even to guess n_D . Finally, it might be argued that social networks, given their perceived neutrality (or even resistance) to control, should have a high n_D , as it is necessary to control most individuals separately to control the whole system. We used the mapping into maximum matching to determine the minimum set of driver nodes (N_D) for the networks in Table 1, the

obtained trend defying our expectations: as a group, gene regulatory networks display high n_D (~ 0.8), indicating that it is necessary to independently control about 80% of nodes to control them fully. In contrast, several social networks are characterized by some of the smallest n_D values, suggesting that a few individuals could in principle control the whole system.

Given the important role hubs (nodes with high degree) have in maintaining the structural integrity of networks against failures and attacks^{31,32}, in spreading phenomena^{32,33} and in synchronization^{8,34}, it is natural to expect that control of the hubs is essential to control a network. To test the validity of this hypothesis, we divided the nodes into three groups of equal size according to their degree, k (low, medium and high). As Fig. 2a, b shows for two canonical network models (Erdős–Rényi^{35,36} and scale-free^{15,37–39}), the fraction of driver nodes is significantly higher among low- k nodes than among the hubs. In Fig. 2c, we plot the mean degree of the driver nodes, $\langle k_D \rangle$, as a function of the mean degree, $\langle k \rangle$, of each network in Table 1 and several network models. In all cases, $\langle k_D \rangle$ is either significantly smaller than or comparable to $\langle k \rangle$, indicating that in both real and model systems the driver nodes tend to avoid the hubs.

To identify the topological features that determine network controllability, we randomized each real network using a full randomization procedure (rand-ER) that turns the network into a directed Erdős–Rényi random network with N and L unchanged. For several

networks there is no correlation between the N_D of the original network and the N_D of its randomized counterpart (Fig. 2d), indicating that full randomization eliminates the topological characteristics that influence controllability. We also applied a degree-preserving randomization^{40,41} (rand-Degree), which keeps the in-degree, k_{in} , and out-degree, k_{out} , of each node unchanged but selects randomly the nodes that link to each other. We find that this procedure does not alter N_D significantly, despite the observed differences in N_D of six orders of magnitude (Fig. 2e). Thus, a system's controllability is to a great extent encoded by the underlying network's degree distribution, $P(k_{in}, k_{out})$, which is our second and most important finding. It indicates that N_D is determined mainly by the number of incoming and outgoing links each node has and is independent of where those links point.

An analytical approach to controllability

The importance of the degree distribution allows us to determine N_D analytically for a network with an arbitrary $P(k_{in}, k_{out})$. Using the cavity method^{42–44}, we derived a set of self-consistent equations (Supplementary Information, section IV) whose input is the degree distribution and whose solution is the average n_D (or N_D) over all network realizations compatible with $P(k_{in}, k_{out})$, which is our third key result. As Fig. 2f shows, the analytically predicted N_D agrees perfectly with $N_D^{\text{rand-Degree}}$ (and hence is in good agreement with the exact value, N_D^{real}), offering an effective analytical tool to study

Table 1 | The characteristics of the real networks analysed in the paper

Type	Name	N	L	n_D^{real}	$n_D^{\text{rand-Degree}}$	$n_D^{\text{rand-ER}}$
Regulatory	TRN-Yeast-1	4,441	12,873	0.965	0.965	0.083
	TRN-Yeast-2	688	1,079	0.821	0.811	0.303
	TRN-EC-1	1,550	3,340	0.891	0.891	0.188
	TRN-EC-2	418	519	0.751	0.752	0.380
	Ownership-USCorp	7,253	6,726	0.820	0.815	0.480
Trust	College student	32	96	0.188	0.173	0.082
	Prison inmate	67	182	0.134	0.144	0.103
	Slashdot	82,168	948,464	0.045	0.278	1.7×10^{-5}
	WikiVote	7,115	103,689	0.666	0.666	1.4×10^{-4}
	Epinions	75,888	508,837	0.549	0.606	0.001
Food web	Ythan	135	601	0.511	0.433	0.016
	Little Rock	183	2,494	0.541	0.200	0.005
	Grassland	88	137	0.523	0.477	0.301
	Seagrass	49	226	0.265	0.199	0.203
Power grid	Texas	4,889	5,855	0.325	0.287	0.396
Metabolic	<i>Escherichia coli</i>	2,275	5,763	0.382	0.218	0.129
	<i>Saccharomyces cerevisiae</i>	1,511	3,833	0.329	0.207	0.130
	<i>Caenorhabditis elegans</i>	1,173	2,864	0.302	0.201	0.144
Electronic circuits	s838	512	819	0.232	0.194	0.293
	s420	252	399	0.234	0.195	0.298
	s208	122	189	0.238	0.199	0.301
Neuronal	<i>Caenorhabditis elegans</i>	297	2,345	0.165	0.098	0.003
Citation	ArXiv-HepTh	27,770	352,807	0.216	0.199	3.6×10^{-5}
	ArXiv-HepPh	34,546	421,578	0.232	0.208	3.0×10^{-5}
World Wide Web	nd.edu	325,729	1,497,134	0.677	0.622	0.012
	stanford.edu	281,903	2,312,497	0.317	0.258	3.0×10^{-4}
	Political blogs	1,224	19,025	0.356	0.285	8.0×10^{-4}
Internet	p2p-1	10,876	39,994	0.552	0.551	0.001
	p2p-2	8,846	31,839	0.578	0.569	0.002
	p2p-3	8,717	31,525	0.577	0.574	0.002
Social communication	UCLonline	1,899	20,296	0.323	0.322	0.706
	Email-epoch	3,188	39,256	0.426	0.332	3.0×10^{-4}
	Cellphone	36,595	91,826	0.204	0.212	0.133
Intra-organizational	Freemans-2	34	830	0.029	0.029	0.029
	Freemans-1	34	695	0.029	0.029	0.029
	Manufacturing	77	2,228	0.013	0.013	0.013
	Consulting	46	879	0.043	0.043	0.022

For each network, we show its type and name; number of nodes (N) and edges (L); and density of driver nodes calculated in the real network (n_D^{real}), after degree-preserved randomization ($n_D^{\text{rand-Degree}}$) and after full randomization ($n_D^{\text{rand-ER}}$). For data sources and references, see Supplementary Information, section VI.

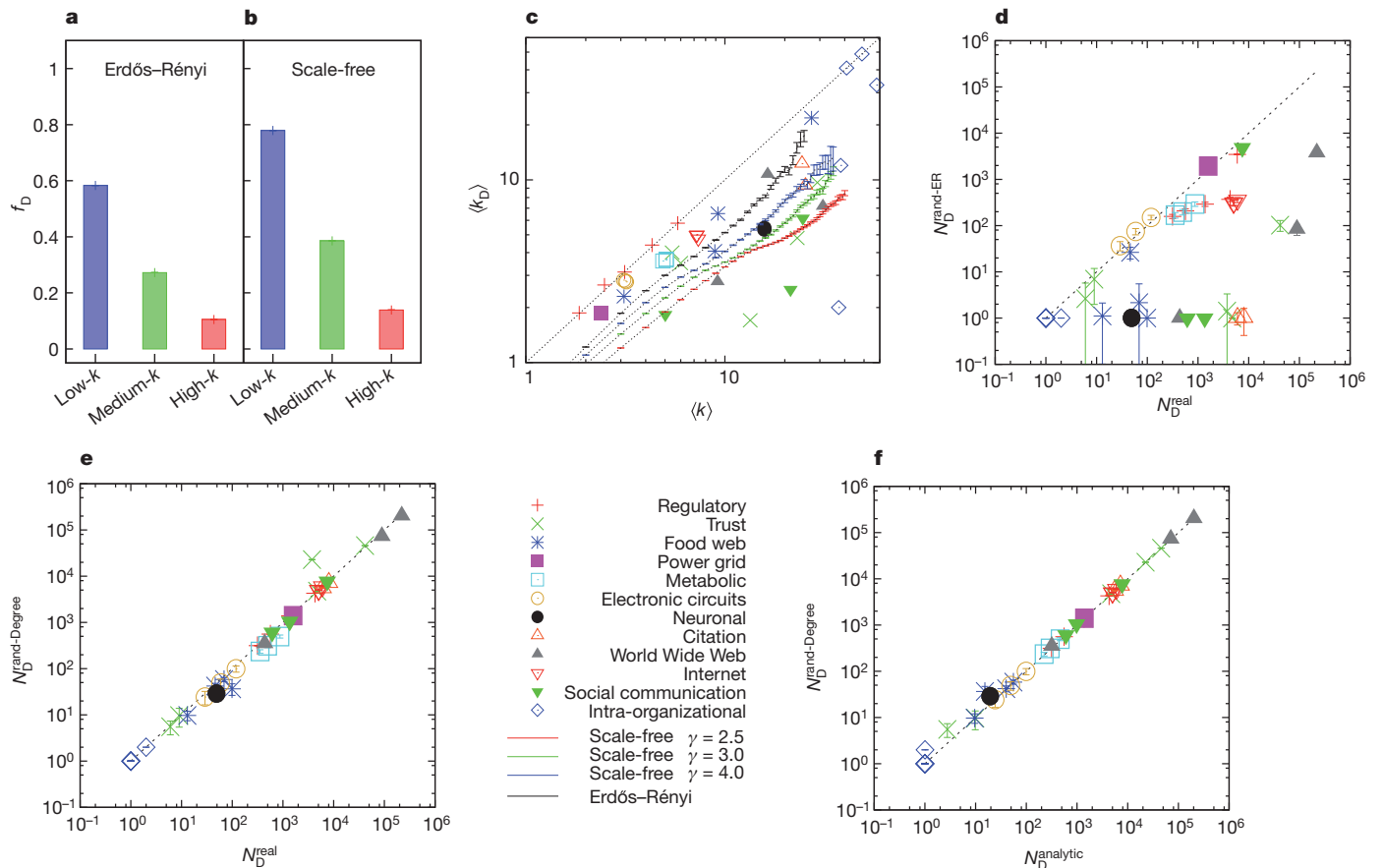


Figure 2 | Characterizing and predicting the driver nodes (N_D). **a, b,** Role of the hubs in model networks. The bars show the fractions of driver nodes, f_D , among the low-, medium- and high-degree nodes in two network models, Erdős-Rényi (**a**) and scale-free (**b**), with $N = 10^4$ and $\langle k \rangle = 3$ ($\gamma = 3$), indicating that the driver nodes tend to avoid the hubs. Both the Erdős-Rényi and the scale-free networks are generated from the static model³⁸ and the results are averaged over 100 realizations. The error bars (s.e.m.), shown in the figure, are smaller than the symbols. **c,** Mean degree of driver nodes compared with the mean degree of all nodes in real and model networks, indicating that in real

systems the hubs are avoided by the driver nodes. **d,** Number of driver nodes, $N_D^{\text{rand-ER}}$, obtained for the fully randomized version of the networks listed in Table 1, compared with the exact value, N_D^{real} . **e,** Number of driver nodes, $N_D^{\text{rand-Degree}}$, obtained for the degree-preserving randomized version of the networks shown in Table 1, compared with N_D^{real} . **f,** The analytically predicted N_D^{analytic} calculated using the cavity method, compared with $N_D^{\text{rand-Degree}}$. In **d–f**, data points and error bars (s.e.m.) were determined from 1,000 realizations of the randomized networks.

the impact of various network parameters on N_D . Although the cavity method does not offer a closed-form solution, we can derive the dependence of n_D on key network parameters in the thermodynamic limit ($N \rightarrow \infty$). We find, for example, that for a directed Erdős-Rényi network n_D decays as

$$n_D \approx e^{-\langle k \rangle / 2} \quad (4)$$

for large $\langle k \rangle$. For scale-free networks with degree exponent $\gamma_{\text{in}} = \gamma_{\text{out}} = \gamma$ in the large- $\langle k \rangle$ limit³⁸, we have

$$n_D \approx \exp \left[-\frac{1}{2} \left(1 - \frac{1}{\gamma - 1} \right) \langle k \rangle \right] \quad (5)$$

which has the same $\langle k \rangle$ dependence as equation (4) in the $\gamma \rightarrow \infty$ limit. Equation (5) predicts that $\gamma_c = 2$ is a critical exponent for the controllability of an infinite scale-free network, as only for $\gamma > \gamma_c$ can we control the full system through a finite subset of nodes (that is, $n_D < 1$). For $\gamma \leq \gamma_c$ in the thermodynamic limit, all nodes must be individually controlled (that is, $n_D = 1$). We note that γ_c is different from $\gamma = 3$, which is the critical exponent for a number of network phenomena driven by the divergence of $\langle k^2 \rangle$, from network robustness to epidemic spreading^{31–33,45}. To check the validity of the analytical predictions, we determined the $\langle k \rangle$ dependence of n_D numerically for both Erdős-Rényi and scale-free networks, confirming the asymptotic exponential dependence of n_D on $\langle k \rangle$, as predicted by equations (4) and (5). Furthermore, the predicted n_D value is in excellent

agreement with the numerical results for $\gamma > 3$ (Fig. 3d, e). Near $\gamma = 2$, however, n_D as predicted by the cavity method deviates from the exact n_D value owing to degree correlations that are prominent at $\gamma_c = 2$ and can be eliminated by imposing a degree cut-off in constructing the scale-free networks^{39,46} (Supplementary Information, section IV.B).

Equation (5) also shows that n_D decreases as γ increases (for fixed $\langle k \rangle$), indicating that n_D is affected by degree heterogeneity, representing the spread between the less connected and the more connected nodes. We defined the degree heterogeneity as $H = \Delta / \langle k \rangle$, where $\Delta = \sum_i \sum_j |k_i - k_j| P(k_i) P(k_j)$ is the average absolute degree difference of all pairs of nodes (i and j) drawn from the degree distribution $P(k)$. The degree heterogeneity is zero ($H = 0$) for networks in which all nodes have the same degree, such as the random regular digraph (Fig. 3a), in which the in- and out-degrees of the nodes are fixed to $\langle k \rangle / 2$ but the nodes are connected randomly. For $\langle k \rangle \geq 2$, this graph always has a perfect matching⁴⁷, which means that a single driver node can control the whole system (Supplementary Information, section IV.B1). The degree heterogeneity increases as we move from the random regular digraph to an Erdős-Rényi network (Fig. 3b) and eventually to scale-free networks with decreasing γ (Fig. 3c). Overall, the fraction of driver nodes, n_D , increases monotonically with H , whether we keep γ (Fig. 3f) or $\langle k \rangle$ (Fig. 3g) constant.

Taking these results together, we find that the denser a network is, the fewer driver nodes are needed to control it, and that small changes in the average degree induce orders-of-magnitude variations in n_D .

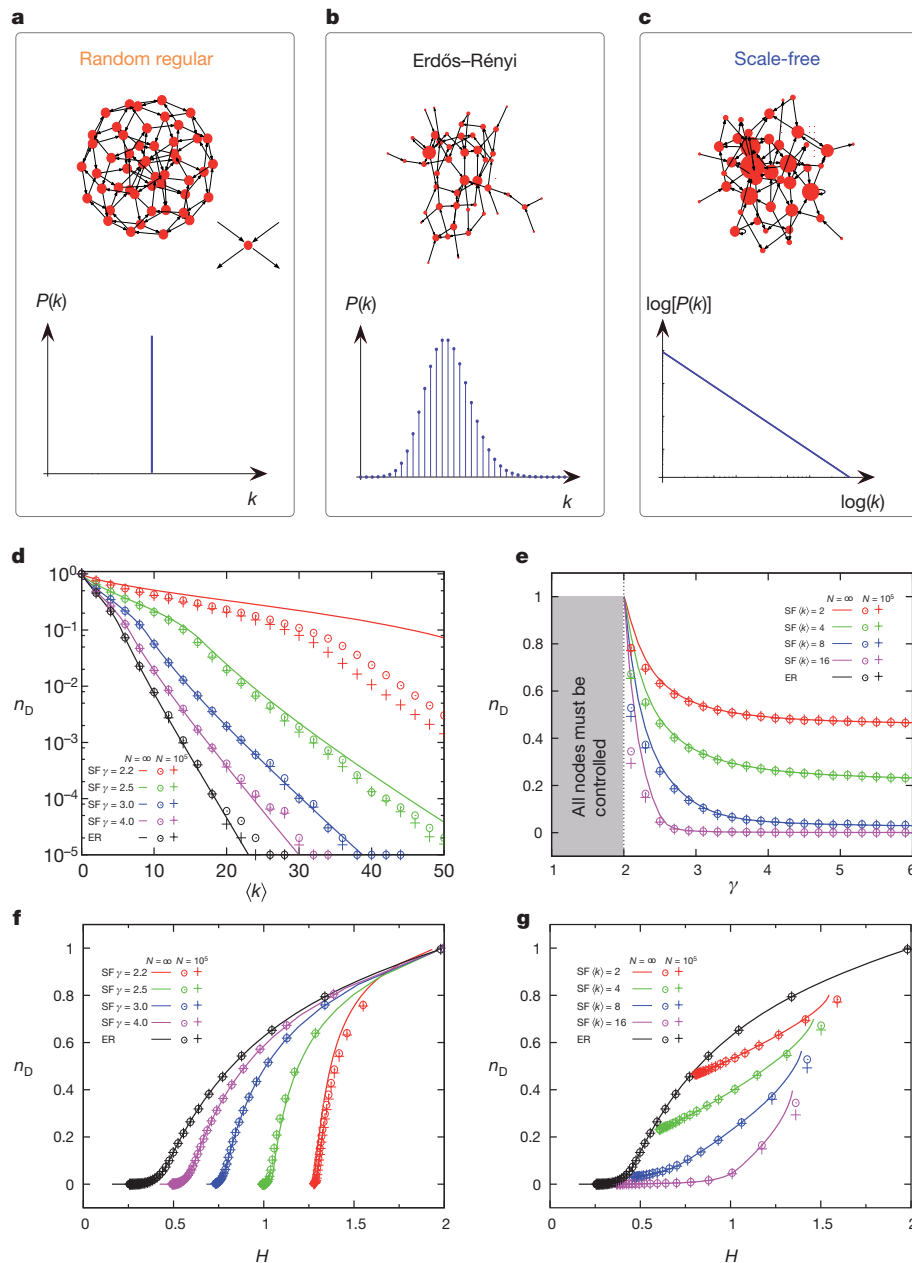


Figure 3 | The impact of network structure on the number of driver nodes. **a–c**, Characteristics of the explored model networks. A random-regular digraph (**a**), shown here for $\langle k \rangle = 4$, is the most degree-homogeneous network as $k_{\text{in}} = k_{\text{out}} = \langle k \rangle / 2$ for all nodes. Erdős-Rényi networks (**b**) have Poisson degree distributions and their degree heterogeneities are determined by $\langle k \rangle$. Scale-free networks (**c**) have power-law degree distributions, yielding large degree heterogeneities. **d**, Driver node density, n_D , as a function of $\langle k \rangle$ for Erdős-Rényi (ER) and scale-free (SF) networks with different values of γ . Both the Erdős-Rényi and the scale-free networks are generated from the static model³⁸ with $N = 10^5$. Lines are analytical results calculated by the cavity method using the expected degree distribution in the $N \rightarrow \infty$ limit. Symbols are calculated for the constructed discrete network: open circles indicate exact results calculated from the maximum matching algorithm, and plus symbols indicate the analytical results of the cavity method using the exact degree sequence of the constructed network. For large $\langle k \rangle$, n_D approaches its lower bound, N^{-1} , that is, a single driver node ($N_D = 1$) in a network of size N . **e**, n_D as a function of γ for scale-free networks with fixed $\langle k \rangle$. For infinite scale-free networks, $n_D \rightarrow 1$ as $\gamma \rightarrow \gamma_c = 2$, that is, it is necessary to control almost all nodes to control the network fully. For finite scale-free networks, n_D reaches its maximum as γ approaches γ_c (Supplementary Information). **f**, n_D as a function of degree heterogeneity, H , for Erdős-Rényi and scale-free networks with fixed γ and variable $\langle k \rangle$. **g**, n_D as a function of H for Erdős-Rényi and scale-free networks for fixed $\langle k \rangle$ and variable γ . As γ increases, the curves converge to the Erdős-Rényi result (black) at the corresponding $\langle k \rangle$ value.

Furthermore, the larger are the differences between node degrees, the more driver nodes are needed to control the system. Overall, networks that are sparse and heterogeneous, which are precisely the characteristics often seen in complex systems like the cell or the Internet^{13–16}, require the most driver nodes, underscoring that such systems are difficult to control.

Robustness of control

To see how robust is our ability to control a network under unavoidable link failure, we classify each link into one of the following three categories (Fig. 1d, g, j): ‘critical’ if in its absence we need to increase the number of driver nodes to maintain full control; ‘redundant’ if it can be removed without affecting the current set of driver nodes; or ‘ordinary’ if it is neither critical nor redundant. Figure 4 shows the densities of critical ($l_c = L_c/L$), redundant ($l_r = L_r/L$) and ordinary ($l_o = L_o/L$) links for each real network, and indicates that most networks have few or no critical links. Most links are ordinary, meaning that they have a role in some control configurations but that the network can still be controlled in their absence.

To understand the factors that determine l_c , l_r and l_o , in Fig. 5a, c, f we show their $\langle k \rangle$ dependence for model systems. The behaviour of l_c is the easiest to understand: for small $\langle k \rangle$, all links are essential for control ($l_c \approx 1$). As $\langle k \rangle$ increases, the network’s redundancy increases, decreasing l_c . The increasing redundancy suggests that the density of redundant links, l_r , should always increase with $\langle k \rangle$, but it does not: it reaches a maximum at a critical value of $\langle k \rangle$, $\langle k \rangle_c$, after which it decays. This non-monotonic behaviour results from the competition of two topologically distinct regions of a network, the core and leaves⁴³. The core represents a compact cluster of nodes left in the network after applying a greedy leaf removal procedure⁴⁸, and leaves are nodes with $k_{\text{in}} = 1$ or $k_{\text{out}} = 1$ before or during leaf removal. The core emerges through a percolation transition (Fig. 5b, d): for $k < \langle k \rangle_c$, $n_{\text{core}} = N_{\text{core}}/N = 0$, so the system consists of leaves only (Fig. 5e). At $\langle k \rangle = \langle k \rangle_c$, a small core emerges, decreasing the number of leaves. For Erdős-Rényi networks, we predict that $\langle k \rangle_c = 2e \approx 5.436564$ in agreement with the numerical result (Fig. 5a, b), a value that coincides with $\langle k \rangle$ where l_r reaches its maximum. Indeed, l_r starts decaying at $\langle k \rangle_c$ because for $\langle k \rangle > \langle k \rangle_c$ the number of distinct maximum

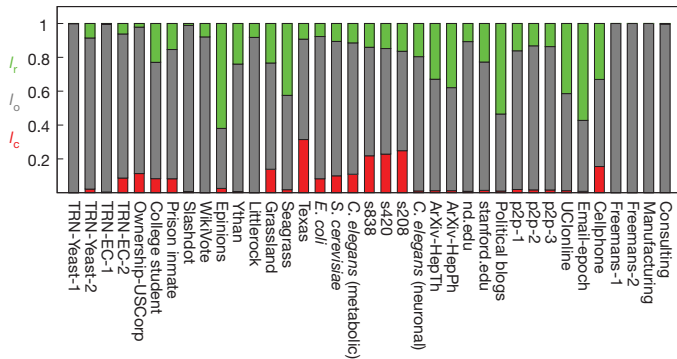


Figure 4 | Link categories for robust control. The fractions of critical (red, l_c), redundant (green, l_r) and ordinary (grey, l_o) links for the real networks named in Table 1. To make controllability robust to link failures, it is sufficient to double only the critical links, formally making each of these links redundant and therefore ensuring that there are no critical links in the system.

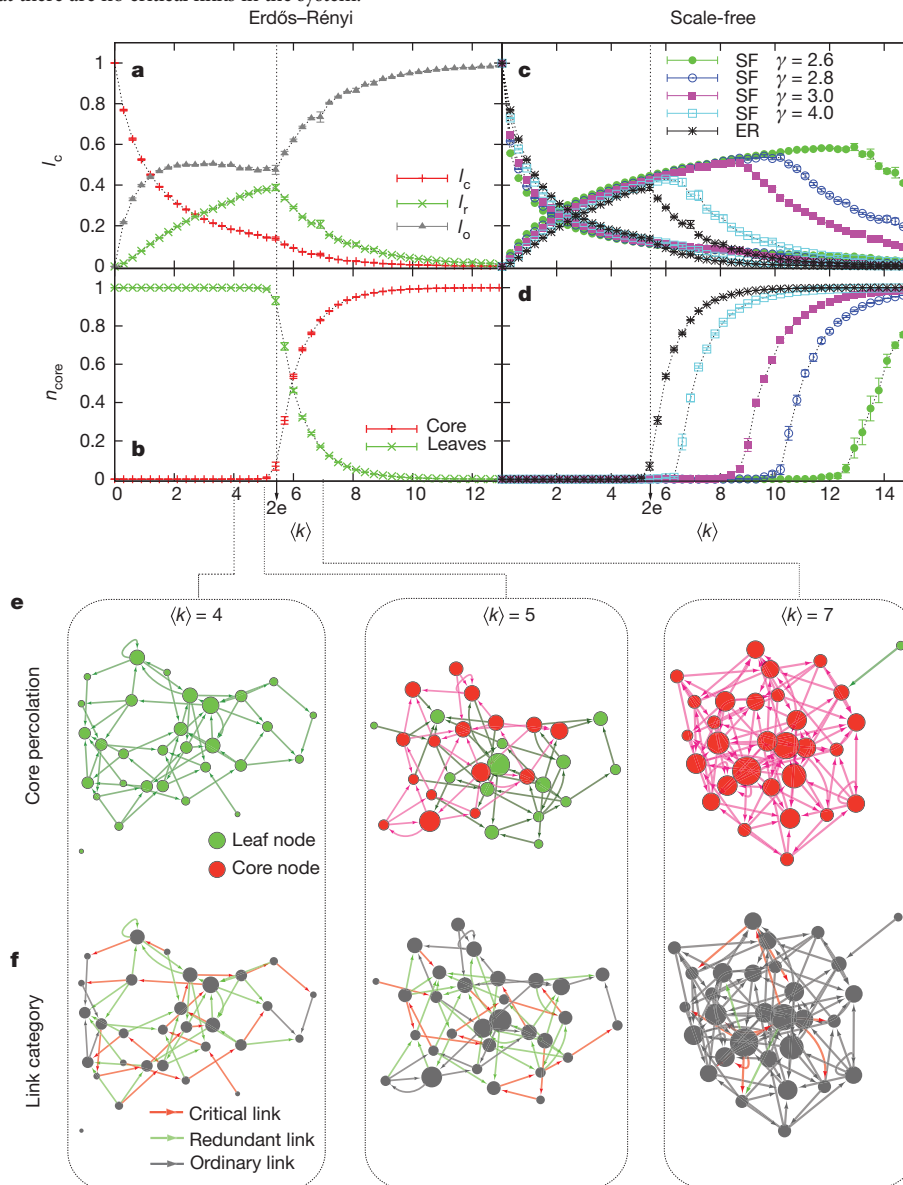


Figure 5 | Control robustness. **a**, Dependence on $\langle k \rangle$ of the fraction of critical (red, l_c), redundant (green, l_r) and ordinary (grey, l_o) links for an Erdős-Rényi network: l_r peaks at $\langle k \rangle = \langle k \rangle_c = 2e$ and the derivative of l_c is discontinuous at $\langle k \rangle = \langle k \rangle_c$. **b**, Core percolation for Erdős-Rényi network occurs at $k = \langle k \rangle_c = 2e$, which explains the l_r peak. **c**, **d**, Same as in **a** and **b** but for scale-free networks. The Erdős-Rényi and scale-free networks³⁸ have $N = 10^4$ and the results are

matchings increases exponentially (Supplementary Information, section IV.C) and, as a result, the chance that a link does not participate in any control configuration decreases. For scale-free networks, we observe the same behaviour, with the caveat that $\langle k \rangle_c$ decreases with γ (Fig. 5c, d).

Discussion and conclusions

Control is a central issue in most complex systems, but because a general theory to explore it in a quantitative fashion has been lacking, little is known about how we can control a weighted, directed network—the configuration most often encountered in real systems. Indeed, applying Kalman's controllability rank condition (equation (3)) to large networks is computationally prohibitive, limiting previous work to a few dozen nodes at most^{17–19}. Here we have developed the tools to address controllability for arbitrary network topologies and sizes. Our key finding, that N_D is determined mainly by the degree

averaged over ten realizations with error bars defined as s.e.m. Dotted lines are only a guide to the eye. **e**, The core (red) and leaves (green) for small Erdős-Rényi networks ($N = 30$) at different $\langle k \rangle$ values ($\langle k \rangle = 4, 5, 7$). Node sizes are proportional to node degrees. **f**, The critical (red), redundant (green) and ordinary (grey) links for the above Erdős-Rényi networks at the corresponding $\langle k \rangle$ values.

distribution, allows us to use the tools of statistical physics to predict N_D from $P(k_{in}, k_{out})$ analytically, offering a general formalism with which to explore the impact of network topology on controllability.

The framework presented here raises a number of questions, answers to which could further deepen our understanding of control in complex environments. For example, although our analytical work focused on uncorrelated networks, the algorithmic method we developed can identify N_D for arbitrary networks, providing a framework in which to address the role of correlations systematically^{40,49,50}. Taken together, our results indicate that many aspects of controllability can be explored exactly and analytically for arbitrary networks if we combine the tools of network science and control theory, opening new avenues to deepening our understanding of complex systems.

Received 18 November 2010; accepted 16 March 2011.

- Kalman, R. E. Mathematical description of linear dynamical systems. *J. Soc. Indust. Appl. Math. Ser. A* **1**, 152–192 (1963).
- Luenberger, D. G. *Introduction to Dynamic Systems: Theory, Models, & Applications* (Wiley, 1979).
- Slotine, J.-J. & Li, W. *Applied Nonlinear Control* (Prentice-Hall, 1991).
- Kelly, F. P., Maulloo, A. K. & Tan, D. K. H. Rate control for communication networks: shadow prices, proportional fairness and stability. *J. Oper. Res. Soc.* **49**, 237–252 (1998).
- Srikant, R. *The Mathematics of Internet Congestion Control* (Birkhäuser, 2004).
- Chiang, M., Low, S. H., Calderbank, A. R. & Doyle, J. C. Layering as optimization decomposition: a mathematical theory of network architectures. *Proc. IEEE* **95**, 255–312 (2007).
- Wang, X. F. & Chen, G. Pinning control of scale-free dynamical networks. *Physica A* **310**, 521–531 (2002).
- Wang, W. & Slotine, J.-J. E. On partial contraction analysis for coupled nonlinear oscillators. *Biol. Cybern.* **92**, 38–53 (2005).
- Sorrentino, F., di Bernardo, M., Garofalo, F. & Chen, G. Controllability of complex networks via pinning. *Phys. Rev. E* **75**, 046103 (2007).
- Yu, W., Chen, G. & Lü, J. On pinning synchronization of complex dynamical networks. *Automatica* **45**, 429–435 (2009).
- Marucci, L. *et al.* How to turn a genetic circuit into a synthetic tunable oscillator, or a bistable switch. *PLoS ONE* **4**, e8083 (2009).
- Strogatz, S. H. Exploring complex networks. *Nature* **410**, 268–276 (2001).
- Dorogovtsev, S. N. & Mendes, J. F. F. *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford Univ. Press, 2003).
- Newman, M., Barabási, A.-L. & Watts, D. J. *The Structure and Dynamics of Networks* (Princeton Univ. Press, 2006).
- Caldarelli, G. *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford Univ. Press, 2007).
- Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
- Tanner, H. G. in *Proc. 43rd IEEE Conf. Decision Contr.* Vol. 3, 2467–2472 (2004).
- Lombardi, A. & Hörnquist, M. Controllability analysis of networks. *Phys. Rev. E* **75**, 56110 (2007).
- Liu, B., Chu, T., Wang, L. & Xie, G. Controllability of a leader–follower dynamic network with switching topology. *IEEE Trans. Automat. Contr.* **53**, 1009–1013 (2008).
- Rahmani, A., Ji, M., Mesbahi, M. & Egerstedt, M. Controllability of multi-agent systems from a graph-theoretic perspective. *SIAM J. Contr. Optim.* **48**, 162–186 (2009).
- Kim, D.-H. & Motter, A. E. Slave nodes and the controllability of metabolic networks. *N. J. Phys.* **11**, 113047 (2009).
- Mesbahi, M. & Egerstedt, M. *Graph Theoretic Methods in Multiagent Networks* (Princeton Univ. Press, 2010).
- Motter, A. E., Gulbahce, N., Almaas, E. & Barabási, A.-L. Predicting synthetic rescues in metabolic networks. *Mol. Syst. Biol.* **4**, 168 (2008).
- Pastor-Satorras, R. & Vespignani, A. *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge Univ. Press, 2004).
- Lezon, T. R., Banavar, J. R., Cieplak, M., Maritan, A. & Fedoroff, N. V. Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc. Natl Acad. Sci. USA* **103**, 19033–19038 (2006).
- Lin, C.-T. Structural controllability. *IEEE Trans. Automat. Contr.* **19**, 201–208 (1974).
- Shields, R. W. & Pearson, J. B. Structural controllability of multi-input linear systems. *IEEE Trans. Automat. Contr.* **21**, 203–212 (1976).
- Lohmiller, W. & Slotine, J.-J. E. On contraction analysis for nonlinear systems. *Automatica* **34**, 683–696 (1998).
- Yu, W., Chen, G., Cao, M. & Kurths, J. Second-order consensus for multiagent systems with directed topologies and nonlinear dynamics. *IEEE Trans. Syst. Man Cybern. B* **40**, 881–891 (2010).
- Hopcroft, J. E. & Karp, R. M. An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM J. Comput.* **2**, 225–231 (1973).
- Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
- Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.* **85**, 4626–4628 (2000).
- Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203 (2001).
- Nishikawa, T., Motter, A. E., Lai, Y.-C. & Hoppensteadt, F. C. Heterogeneity in oscillator networks: are smaller worlds easier to synchronize? *Phys. Rev. Lett.* **91**, 014101 (2003).
- Erdős, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–60 (1960).
- Bollobás, B. *Random Graphs* (Cambridge Univ. Press, 2001).
- Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- Goh, K.-I., Kahng, B. & Kim, D. Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.* **87**, 278701 (2001).
- Chung, F. & Lu, L. Connected component in random graphs with given expected degree sequences. *Ann. Combin.* **6**, 125–145 (2002).
- Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).
- Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
- Mézard, M. & Parisi, G. The Bethe lattice spin glass revisited. *Eur. Phys. J. B* **20**, 217–233 (2001).
- Zdeborová, L. & Mézard, M. The number of matchings in random graphs. *J. Stat. Mech.* **05**, 05003 (2006).
- Zhou, H. & Ou-Yang, Z.-c. Maximum matching on random graphs. Preprint at (<http://arxiv.org/abs/cond-mat/0309348>) (2003).
- Callaway, D. S., Newman, M. E. J., Strogatz, S. H. & Watts, D. J. Network robustness and fragility: percolation on random graphs. *Phys. Rev. Lett.* **85**, 5468–5471 (2000).
- Boguñá, M., Pastor-Satorras, R. & Vespignani, A. Cut-offs and finite size effects in scale-free networks. *Eur. Phys. J. B* **38**, 205–209 (2004).
- Lovász, L. & Plummer, M. D. *Matching Theory* (American Mathematical Society, 2009).
- Bauer, M. & Golinelli, O. Core percolation in random graphs: a critical phenomena analysis. *Eur. Phys. J. B* **24**, 339–352 (2001).
- Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).
- Pastor-Satorras, R., Vázquez, A. & Vespignani, A. Dynamical and correlation properties of the Internet. *Phys. Rev. Lett.* **87**, 258701 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Song, G. Bianconi, H. Zhou, L. Vepstas, N. Gulbahce, H. Jeong, Y.-Y. Ahn, B. Barzel, N. Blumm, D. Wang, Z. Qu and Y. Li for discussions. This work was supported by the Network Science Collaborative Technology Alliance sponsored by the US Army Research Laboratory under Agreement Number W911NF-09-2-0053; the Office of Naval Research under Agreement Number N000141010968; the Defense Threat Reduction Agency awards WMD BRBA07-J-2-0035 and BRBA08-Per4-C-2-0033; and the James S. McDonnell Foundation 21st Century Initiative in Studying Complex Systems.

Author Contributions All authors designed and did the research. Y.-Y.L. analysed the empirical data and did the analytical and numerical calculations. A.-L.B. was the lead writer of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.-L.B. (alb@neu.edu).

Enterotypes of the human gut microbiome

Manimozhiyan Arumugam^{1*}, Jeroen Raes^{1,2*}, Eric Pelletier^{3,4,5}, Denis Le Paslier^{3,4,5}, Takuji Yamada¹, Daniel R. Mende¹, Gabriel R. Fernandes^{1,6}, Julien Tap^{1,7}, Thomas Bruls^{3,4,5}, Jean-Michel Batto⁷, Marcelo Bertalan⁸, Natalia Borruel⁹, Francesc Casellas⁹, Leyden Fernandez¹⁰, Laurent Gautier⁸, Torben Hansen^{11,12}, Masahira Hattori¹³, Tetsuya Hayashi¹⁴, Michiel Kleerebezem¹⁵, Ken Kurokawa¹⁶, Marion Leclerc⁷, Florence Levenez⁷, Chaysavanh Manichanh⁹, H. Bjørn Nielsen⁸, Trine Nielsen¹¹, Nicolas Pons⁷, Julie Poulain³, Junjie Qin¹⁷, Thomas Sicheritz-Ponten^{8,18}, Sebastian Tims¹⁵, David Torrents^{10,19}, Edgardo Ugarte³, Erwin G. Zoetendal¹⁵, Jun Wang^{17,20}, Francisco Guarner⁹, Oluf Pedersen^{11,21,22,23}, Willem M. de Vos^{15,24}, Søren Brunak⁸, Joel Doré⁷, MetaHIT Consortium†, Jean Weissenbach^{3,4,5}, S. Dusko Ehrlich⁷ & Peer Bork^{1,25}

Our knowledge of species and functional composition of the human gut microbiome is rapidly increasing, but it is still based on very few cohorts and little is known about variation across the world. By combining 22 newly sequenced faecal metagenomes of individuals from four countries with previously published data sets, here we identify three robust clusters (referred to as enterotypes hereafter) that are not nation or continent specific. We also confirmed the enterotypes in two published, larger cohorts, indicating that intestinal microbiota variation is generally stratified, not continuous. This indicates further the existence of a limited number of well-balanced host–microbial symbiotic states that might respond differently to diet and drug intake. The enterotypes are mostly driven by species composition, but abundant molecular functions are not necessarily provided by abundant species, highlighting the importance of a functional analysis to understand microbial communities. Although individual host properties such as body mass index, age, or gender cannot explain the observed enterotypes, data-driven marker genes or functional modules can be identified for each of these host properties. For example, twelve genes significantly correlate with age and three functional modules with the body mass index, hinting at a diagnostic potential of microbial markers.

Various studies of the human intestinal tract microbiome based on the 16S ribosomal-RNA-encoding gene reported species diversity within and between individuals^{1–3}, and the first metagenomics studies characterized the functional repertoire of the microbiomes of several American^{4,5} and Japanese⁶ individuals. Although a general consensus about the phylum level composition in the human gut is emerging^{1,3,7}, the variation in species composition^{1,2} and gene pools^{5,8} within the human population is less clear. Furthermore, it is unknown whether inter-individual variation manifests itself as a continuum of different community compositions or whether individual gut microbiota congregate around preferred, balanced and stable community compositions that can be classified. Studying such questions is complicated by the complexity of sampling, DNA preparation, processing, sequencing and analysis protocols⁹ as well as by varying physiological, nutritional and environmental conditions. To analyse the feasibility of comparative metagenomics of the human gut across cohorts and protocols and to obtain first insights into commonalities and differences between gut microbiomes across different populations, we Sanger-sequenced 22 European metagenomes from Danish, French, Italian and Spanish individuals that were selected for diversity (Supplementary Notes section 1), and combined them with existing Sanger

(13 Japanese⁶, 2 American⁴) and pyrosequencing (2 American⁵) gut data sets—totalling 39 individuals.

Global variation of human gut metagenomes

The vast majority of sequences in the newly sequenced 22 European samples belong to bacteria—only 0.14% of the reads could be classified as human contamination, all other eukaryotes together only comprised 0.5%, archaea 0.8% and viruses up to 5.8% (see Supplementary Notes section 2.1 for details).

To investigate the phylogenetic composition of the 39 samples from 6 nationalities, we mapped metagenomic reads, using DNA sequence homology, to 1,511 reference genomes (Supplementary Table 3) including 379 publicly available human microbiome genomes generated through the National Institutes of Health (NIH) Human Microbiome Project¹⁰ and the European MetaHIT consortium¹¹ (Supplementary Methods section 4.1). To consistently estimate the functional composition of the samples, we annotated the predicted genes from the metagenomes using eggNOG¹² orthologous groups (Supplementary Methods section 6.2). We ensured that comparative analysis using these procedures was not biased by data-set origin, sample preparation, sequencing technology and quality filtering (see Supplementary Notes section 1).

¹European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. ²VIB—Vrije Universiteit Brussel, 1050 Brussels, Belgium. ³Commissariat à l'Energie Atomique, Genoscope, 91000 Evry, France. ⁴Centre National de la Recherche Scientifique, UMR8030, 91000 Evry, France. ⁵Université d'Evry Val d'Essonne 91000 Evry, France. ⁶Department of Biochemistry and Immunology, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, Minas Gerais, Brazil. ⁷Institut National de la Recherche Agronomique, 78350 Jouy en Josas, France. ⁸Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Lyngby, Denmark. ⁹Digestive System Research Unit, University Hospital Vall d'Hebron, Ciberehd, 08035 Barcelona, Spain.

¹⁰Barcelona Supercomputing Center, Jordi Girona 31, 08034 Barcelona, Spain. ¹¹Marie Krogh Center for Metabolic Research, Section of Metabolic Genetics, Faculty of Health Sciences, University of Copenhagen, DK-2100 Copenhagen, Denmark. ¹²Faculty of Health Sciences, University of Southern Denmark, DK-5000 Odense, Denmark. ¹³Computational Biology Laboratory Bld, The University of Tokyo Kashiwa Campus, Kashiwa-no-ha 5-1-5, Kashiwa, Chiba, 277-8561, Japan. ¹⁴Division of Bioenvironmental Science, Frontier Science Research Center, University of Miyazaki, 5200 Kiyotake, Miyazaki 889-1692, Japan. ¹⁵Laboratory of Microbiology, Wageningen University, 6710BA Ede, The Netherlands. ¹⁶Tokyo Institute of Technology, Graduate School of Bioscience and Biotechnology, Department of Biological Information, 4259 Nagatsuta-cho, Midori-ku, Yokohama-shi, Kanagawa Pref. 226-8501, Japan. ¹⁷BGI-Shenzhen, Shenzhen 518083, China. ¹⁸Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Lyngby, Denmark. ¹⁹Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain. ²⁰Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark. ²¹Institute of Biomedical Science, Faculty of Health Sciences, University of Copenhagen, DK-2200 Copenhagen, Denmark.

²²Hagedorn Research Institute, DK-2820 Gentofte, Denmark. ²³Faculty of Health Sciences, University of Aarhus, DK-8000 Aarhus, Denmark. ²⁴University of Helsinki, FI-00014 Helsinki, Finland. ²⁵Max Delbrück Centre for Molecular Medicine, D-13092 Berlin, Germany.

*These authors contributed equally to this work.

†Lists of authors and affiliations appear at the end of the paper.

species, which could shed light on their survival strategies in the human gut. In the samples analysed here, the most abundant molecular functions generally trace back to the most dominant species. However, we identified some abundant orthologous groups that are contributed to primarily by low-abundance genera (see Supplementary Fig. 2, Supplementary Table 6 and Supplementary Notes section 3). For example, low-abundance *Escherichia* contribute over 90% of two abundant proteins associated with bacterial pilus assembly, FimA (COG3539)

and PapC (COG3188), found in one individual (IT-AD-5). Pili enable the microbes to colonize the epithelium of specific host organs; they help microbes to stay longer in the human intestinal tract by binding to human mucus or mannose sugars present on intestinal surface structures¹⁸. They are also key components in the transfer of plasmids between bacteria through conjugation, often leading to exchange of protective functions such as antibiotic resistance¹⁸. Pili can thus provide multiple benefits to these low-abundance microbes in their efforts

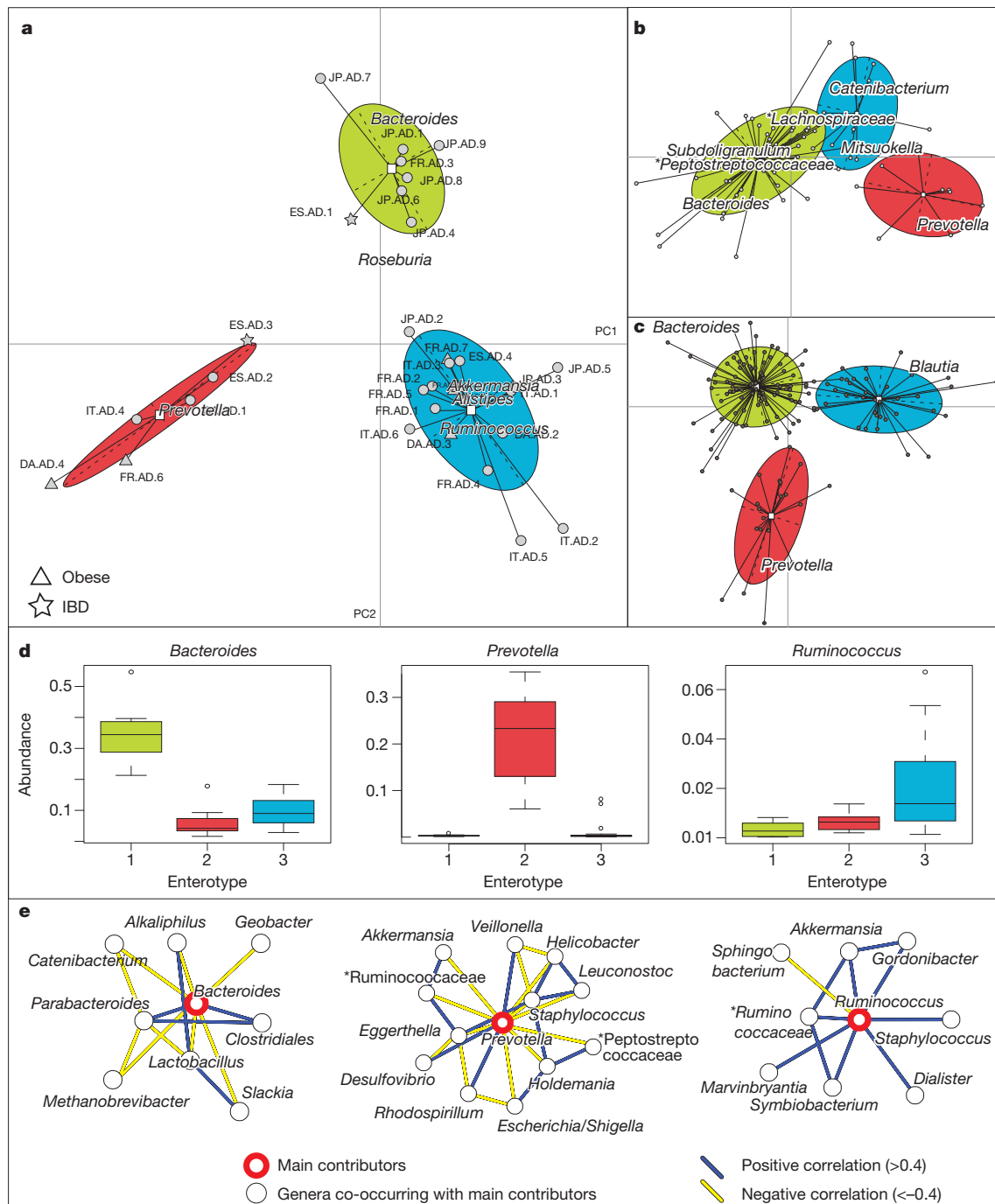


Figure 2 | Phylogenetic differences between enterotypes. **a–c**, Between-class analysis, which visualizes results from PCA and clustering, of the genus compositions of 33 Sanger metagenomes estimated by mapping the metagenome reads to 1,511 reference genome sequences using an 85% similarity threshold (**a**), Danish subset containing 85 metagenomes from a published Illumina data set⁸ (**b**) and 154 pyrosequencing-based 16S sequences⁵ (**c**) reveal three robust clusters that we call enterotypes. IBD, inflammatory bowel disease. Two principal components are plotted using the ade4 package in

R with each sample represented by a filled circle. The centre of gravity for each cluster is marked by a rectangle and the coloured ellipse covers 67% of the samples belonging to the cluster. IBD, inflammatory bowel disease. **d**, Abundances of the main contributors of each enterotype from the Sanger metagenomes. See Fig. 1 for definition of box plot. **e**, Co-occurrence networks of the three enterotypes from the Sanger metagenomes. Unclassified genera under a higher rank are marked by asterisks in **b** and **e**.

to survive and persist in the human gut. This example illustrates that abundant species or genera cannot reveal the entire functional complexity of the gut microbiota. More reference genomes will facilitate better taxonomic assignment from samples and thus the detection of more low-abundance species. However, there is not much room for as yet undetected, abundant genera. Even with our limited genus assignment rate of 52.8% of all reads, we estimate that we miss another 30.7% of the already classified genera owing to our strict assignment criteria (Supplementary Fig. 1); that is, only 16.5% of all reads are likely to belong to hitherto unknown genera.

Detection of enterotypes, cross-national clusters

To get an overview of species variation we used phylogenetic profile similarities obtained by mapping metagenomic reads to the 1,511 reference genomes (Fig. 2a; see Supplementary Methods section 4.1). We excluded the two American Sanger-sequenced samples⁴ from further analysis because of an unusual, very low fraction of Bacteroidetes and suspected technical artefacts¹⁹. Multidimensional cluster analysis and principal component analysis (PCA) revealed that the remaining 33 samples formed three distinct clusters that we designate as enterotypes (see Supplementary Notes section 4.1, Supplementary Fig. 3a and Supplementary Table 8). Each of these three enterotypes are identifiable by the variation in the levels of one of three genera: *Bacteroides* (enterotype 1), *Prevotella* (enterotype 2) and *Ruminococcus* (enterotype 3) (Fig. 2a, d), which was reproduced using independent array-based HITChip²⁰ data in a subset of 22 European samples (Supplementary Fig. 4 and Supplementary Notes section 4.5). The same analysis on two larger published gut microbiome data sets of different origins (16S pyrosequencing data from 154 American individuals⁵ and Illumina-based metagenomics data from 85 Danish individuals⁸; Supplementary Methods section 5) shows that these data sets could also be represented best by three clusters (Supplementary Fig. 3b, c and Supplementary Tables 9, 10). Two of these are also driven by *Bacteroides* and *Prevotella*, whereas the third cluster is mostly driven by related groups of the order Clostridiales, *Blautia* and unclassified Lachnospiraceae in the 16S rDNA and Illumina data, respectively (Fig. 2b, c). This can be explained by a different reference data set in the instance of the 16S rDNA data, different mapping behaviour of short reads in the case of the Illumina data or current taxonomic uncertainties in the Lachnospiraceae and Ruminococcaceae clades (see Supplementary Notes section 4.2). The differences might also hint at community subpopulations within this enterotype, which might only be detectable with substantially more samples. Correlation analysis of the Sanger data revealed that abundances of each of the three discriminating genera strongly correlate (that is, they co-occur or avoid each other) with those of other genera (Fig. 2d; see Supplementary Methods section 11), indicating that the enterotypes are in fact driven by groups of species that together contribute to the preferred community compositions.

We demonstrate further the robustness of the enterotypes using two distinct statistical concepts. First, we used the silhouette coefficient²¹ to validate that the three clusters are superior to clusterings obtained from various randomizations of the genus profile data, indicating a potential role for the interactions between co-occurring genera (see Supplementary Fig. 5 and Supplementary Notes section 4.3). Second, we used supervised learning and cross-validation to establish that these clusters have non-random characteristics that can be modelled and subsequently used to classify new samples (learning on clusters from randomized genus profiles led to considerably worse classification performance; see Supplementary Fig. 6 and Supplementary Notes section 4.4). These consistent results indicate that enterotypes will be identifiable in human gut metagenomes also from larger cohorts.

We then clustered the 33 samples using a purely functional metric: the abundance of the assigned orthologous groups (Fig. 3a). Remarkably, this clustering also showed a similar grouping of the samples with only minor differences (five samples placed in different clusters compared

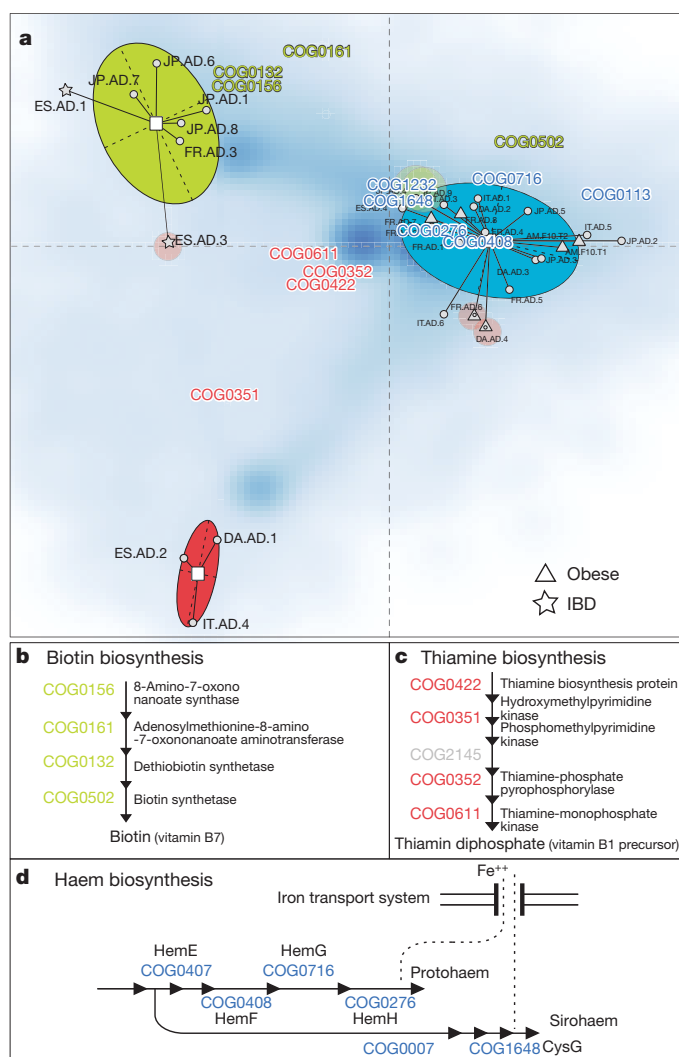


Figure 3 | Functional differences between enterotypes. **a**, Between-class analysis (see Fig. 2) of orthologous group abundances showing only minor disagreements with enterotypes (unfilled circles indicate the differing samples). The blue cloud represents the local density estimated from the coordinates of orthologous groups; positions of selected orthologous groups are highlighted. **b**, Four enzymes in the biotin biosynthesis pathway (COG0132, COG0156, COG0161 and COG0502) are overrepresented in enterotype 1. **c**, Four enzymes in the thiamine biosynthesis pathway (COG0422, COG0351, COG0352 and COG0611) are overrepresented in enterotype 2. **d**, Six enzymes in the haem biosynthesis pathway (COG0007, COG0276, COG407, COG0408, COG0716 and COG1648) are overrepresented in enterotype 3.

to Fig. 2a), indicating that function and species composition roughly coincide with some exceptions such as Spanish sample ES-AD-3, whose genus composition belongs to enterotype 2 whereas its functional composition is similar to members of enterotype 1. This individual has high levels of phage-related genes compared to the other samples (see Supplementary Fig. 7), hinting at partial temporal variability and dynamics of the microbiota, and perhaps indicating phage or virus bursts.

The robustness and predictability of the enterotypes in different cohorts and at multiple phylogenetic and functional levels indicates that they are the result of well-balanced, defined microbial community compositions of which only a limited number exist across individuals. These enterotypes are not as sharply delimited as, for example, human blood groups; they are, in contrast, densely populated areas in a multi-dimensional space of community composition. They are nevertheless likely to characterize individuals, in line with previous reports that gut

microbiota are quite stable in individuals and can even be restored after perturbation^{22–25}.

Variation between enterotypes

To determine the phylogenetic and functional basis of the enterotypes, we investigated in detail their differences in composition at the phylum, genus, gene and pathway level as well as correlations in abundance of co-occurring genera (Figs 2, 3; also see Supplementary Methods sections 10, 11 and 12). Enterotype 1, containing eight samples, is enriched in *Bacteroides* ($P < 0.01$; Supplementary Fig. 8), which co-occurs, for example, with *Parabacteroides* (see Supplementary Table 11 for enriched genera and Fig. 2e for correlation networks of co-occurring genera in each enterotype). The drivers of this enterotype seem to derive energy primarily from carbohydrates and proteins through fermentation, as these closely related genera have a very broad saccharolytic potential²⁶ and because genes encoding enzymes involved in the degradation of these substrates (galactosidases, hexosaminidases, proteases) along with glycolysis and pentose phosphate pathways are enriched in this enterotype (see Supplementary Tables 12, 13). Enterotype 2 contains six samples and is enriched in *Prevotella* ($P < 0.01$; Supplementary Fig. 9) and the co-occurring *Desulfovibrio*, which can act in synergy to degrade mucin glycoproteins present in the mucosal layer of the gut: *Prevotella* is a known mucin-degrader and *Desulfovibrio* may enhance the rate-limiting mucin desulphation step by removing the sulphate²⁷. Enterotype 3 is the most frequent and is enriched in *Ruminococcus* ($P < 0.01$; Supplementary Fig. 10) as well as co-occurring *Akkermansia*, both known to comprise species able to degrade mucins²⁸. It is also enriched in membrane transporters, mostly of sugars, indicating the efficient binding of mucin and its subsequent hydrolysis as well as uptake of the resulting simple sugars by these genera. The enriched genera indicate that enterotypes use different routes to generate energy from fermentable substrates available in the colon, reminiscent of a potential specialization in ecological niches or guilds. In addition to the conversion of complex carbohydrates into absorbable substrates, the gut microbiota is also beneficial to the human host by producing vitamins. Although all the vitamin metabolism pathways are represented in all samples, enterotypes 1 and 2 were enriched in biosynthesis of different vitamins: biotin (Fig. 3b), riboflavin, pantothenate and ascorbate in the former, and thiamine (Fig. 3c) and folate in the latter. These phylogenetic and functional differences among enterotypes thus reflect different combinations of microbial trophic chains with a probable impact on synergistic interrelations with the human hosts.

Functional biomarkers for host properties

Enterotypes do not seem to differ in functional richness (Supplementary Fig. 11), and virtually none of several measured host properties, namely nationality, gender, age or body mass index (BMI), significantly correlates with the enterotypes (with the exception of enterotype 1, which is enriched in Japanese individuals). However, some strong correlations do occur between host properties and particular functions, at the genes or module level (a module is a part of a pathway that is functionally tightly interconnected; see Supplementary Methods sections 6, 13 and Supplementary Notes section 6). The only significant correlation between a host property and a taxonomic group is a negative one between age and the abundance of an unknown Clostridiales genus ($P < 0.02$) containing three obligate anaerobes (Supplementary Fig. 12a; see Supplementary Notes section 6.2). It should be noted that age is not constant across the nationalities (in our data set, Italians are relatively old and Japanese young), but that individuals did not stratify by nationality, indicating that this is not a confounding factor. Our data did not reveal any correlation between BMI and the Firmicutes/Bacteroidetes ratio and we thus cannot contribute to the ongoing debate on the relationship between this ratio and obesity^{29,30}.

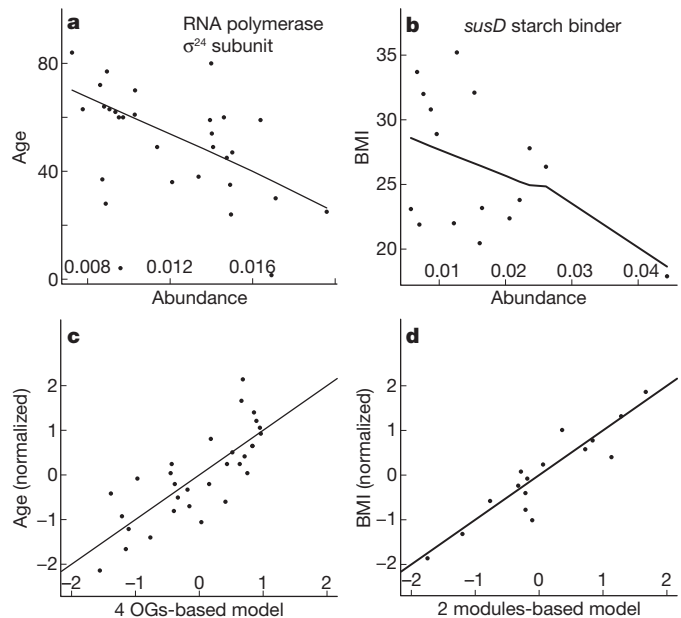


Figure 4 | Correlations with host properties. **a**, Pairwise correlation of RNA polymerase facultative σ^{24} subunit (COG1595) with age ($P = 0.03$, $\rho = -0.59$). **b**, Pairwise correlation of *SusD*, a family of proteins that bind glycan molecules before they are transported into the cell, and BMI ($P = 0.27$, $\rho = -0.29$, weak correlation). **c**, Multiple orthologous groups (OGs) (COG0085, COG0086, COG0438 and COG0739; see Supplementary Table 18) significantly correlating with age when combined into a linear model (see Supplementary Methods section 13 and ref. 40 for details; $P = 2.75 \times 10^{-5}$, adjusted $R^2 = 0.57$). **d**, Two modules, ATPase complex and ectoine biosynthesis (M00051), significantly correlating with BMI when combined into a linear model ($P = 6.786 \times 10^{-6}$, adjusted $R^2 = 0.82$).

In contrast to the minor phylogenetic signal, we found several significant functional correlations with each of the host properties studied (after correcting for multiple testing to avoid artefacts; see Supplementary Methods section 13), indicating that metagenomics-derived functional biomarkers might be more robust than phylogenetic ones. For example, the abundance of ten orthologous groups varies more between than within nationalities (Supplementary Table 14), although overall, the functional composition in total was remarkably similar among the nations (also with respect to the functional core; see Supplementary Fig. 13). For gender, we find five functional modules and one orthologous group that significantly correlate ($P < 0.05$; for example, enriched aspartate biosynthesis modules in males; see Supplementary Table 16). In addition, twelve orthologous groups significantly correlate with age (Supplementary Table 17). For instance, starch degradation enzymes such as glycosidases and glucan phosphorylases increase with age (which could be a reaction to decreased efficiency of host breakdown of dietary carbohydrates with age³¹) and so does the *secA* preprotein translocase (Supplementary Fig. 14). Conversely, an orthologous group coding for the facultative σ^{24} subunit of RNA polymerase, which drives expression under various stress responses and is linked to intestinal survival³², decreases with age (Fig. 4a). One explanation for this could be the reduced need for stress response in the gut due to the age-associated decline in host immune response³³ (immunosenescence). Our analyses also identified three marker modules that correlate strongly with the hosts' BMI (Supplementary Table 19 and Supplementary Fig. 14), two of which are ATPase complexes, supporting the link found between the gut microbiota's capacity for energy harvest and obesity in the host³⁴. Interestingly, functional markers found by a data-driven approach (derived from the metagenomes without previous knowledge) gave much stronger correlations than genes for which a link would be expected (for example, *susC/susD*, involved in starch utilization²⁶;

Fig. 4b). Linear models combining the abundance of only a few functional modules correlate even better with host properties (Fig. 4c, d). It should be noted that given the possibility of many confounding variables owing to the heterogeneity and size of our cohort, these observations will need to be substantiated using larger, independent cohorts in the future. Furthermore, patterns in metagenomics data can (partly) reflect indirect factors⁹ such as genome size³⁵ (the smaller the average genome size of a sample, the higher the relative fraction of single copy genes therein), which, however, does not matter for diagnostics.

Although individual host properties do not explain the enterotypes, the latter might be driven by a complex mixture of functional properties, by host immune modulation or by hitherto unexplored physiological conditions such as transit time or pH of luminal contents. Furthermore, the three major enterotypes could be triggered by the three distinct pathways for hydrogen disposal³⁶ (Supplementary Notes section 6.4). Indeed, despite their low abundance, *Methanobrevibacter* (a methanogen) and *Desulfovibrio* (a known sulphate-reducer) are enriched in enterotypes 3 and 1, respectively.

Taken together, we have demonstrated the existence of enterotypes in the human gut microbiome and have identified three of them that vary in species and functional composition using data that spans several nations and continents. As our current data do not reveal which environmental or even genetic factors are causing the clustering, and as faecal samples are not representative of the entire intestine, we anticipate that the enterotypes introduced here will be refined with deeper and broader analysis of individuals' microbiomes. Presumably, enterotypes are not limited to humans but also occur in animals. Their future investigation might well reveal novel facets of human and animal symbiotic biology and lead to the discovery of those microbial properties correlated with the health status of individuals. We anticipate that they might allow classification of human groups that respond differently to diet or drug intake. Enterotypes appear complex, are probably not driven by nutritional habits and cannot simply be explained by host properties such as age or BMI, although there are functional markers such as genes or modules that correlate remarkably well with individual features. The latter might be utilizable for diagnostic and perhaps even prognostic tools for numerous human disorders, for instance colorectal cancer and obesity-linked co-morbidities such as metabolic syndrome, diabetes and cardiovascular pathologies.

METHODS SUMMARY

Sample collection. Human faecal samples from European individuals were collected and frozen immediately, and DNA was purified as described previously³⁷. Sequencing was carried out by Sanger-sequencing random shotgun DNA libraries of 3 kb using standard protocols established at Genoscope. For sequence processing, cloning vector, sequencing primers and low-quality bases were end-trimmed from raw Sanger reads, and possible human DNA sequences were removed. Reads were processed by the SMASH comparative metagenomics pipeline³⁸ for assembly and gene prediction.

Informed consent was obtained from the 22 European subjects. Sample collection and experiments were approved by the following ethics committees: MetaHIT (Danish), ethical committee of the Capital Region of Denmark; MetaHIT (Spanish), CEIC, Hospital Vall d'Hebron; MicroObes, Ethical Committee for Studies with Human Subjects of Cochin Hospital in Paris, France; MicroAge, Joint Ethical Committee of the University of Camerino.

Phylogenetic annotation. Phylogenetic annotation of samples was performed by (1) aligning reads (Sanger/Illumina) against a database of 1,511 reference genomes (listed in Supplementary Table 3); or (2) classifying 16S rDNA reads using RDP classifier³⁹. Genus and phylum abundance was estimated after normalizing for genome size for the former, and for 16S gene copy number for the latter.

Functional annotation. Genes were functionally annotated using BLASTP against eggNOG (v2) and KEGG (v50) databases. Protein abundances were estimated after normalizing for protein length. Functional abundance profiles at eggNOG, KEGG orthologous group, functional module and pathway level were created.

Clustering and classification. Samples were clustered using Jensen-Shannon distance and partitioning around medoid (PAM) clustering. Optimal number of clusters was estimated using the Calinski-Harabasz (CH) index. We used the silhouette validation technique for assessing the robustness of clusters. Additionally, within a cross-validation scheme, we trained predictive decision tree models on clusters

obtained using the same clustering method and evaluated the classification of hold-out samples by accuracy, average precision and average precision gain.

Statistics. Correlations between metadata and feature abundances were computed as described previously⁴⁰, based on multiple-testing corrected pairwise Spearman correlation analysis and stepwise regression for multi-feature model building. For categorical metadata and enterotype comparisons, samples were pooled into bins (male/female, obese/lean, one enterotype/rest, specific nationality/rest etc) and significant features were identified using Fisher's exact test with multiple testing correction of *P* values.

Received 12 March 2010; accepted 18 December 2010.

Published online 20 April 2011.

- Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638 (2005).
- Hayashi, H., Sakamoto, M. & Benno, Y. Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. *Microbiol. Immunol.* **46**, 535–548 (2002).
- Lay, C. *et al.* Colonic microbiota signatures across five northern European countries. *Appl. Environ. Microbiol.* **71**, 4153–4155 (2005).
- Gill, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355–1359 (2006).
- Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
- Kurokawa, K. *et al.* Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* **14**, 169–181 (2007).
- Zoetendal, E. G., Rajilic-Stojanovic, M. & de Vos, W. M. High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut* **57**, 1605–1615 (2008).
- Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
- Raes, J. & Bork, P. Molecular eco-systems biology: towards an understanding of community function. *Nature Rev. Microbiol.* **6**, 693–699 (2008).
- Nelson, K. E. *et al.* A catalog of reference genomes from the human microbiome. *Science* **328**, 994–999 (2010).
- MetaHIT Consortium. *MetaHIT Draft Bacterial Genomes at the Sanger Institute*. (<http://www.sanger.ac.uk/resources/downloads/bacteria/metahit/>) (9 July 2010).
- Muller, J. *et al.* eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res.* **38**, D190–D195 (2010).
- Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. Development of the human infant intestinal microbiota. *PLoS Biol.* **5**, e177 (2007).
- Tap, J. *et al.* Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* **11**, 2574–2584 (2009).
- Jensen, L. J. *et al.* STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37**, D412–D416 (2009).
- Dethlefsen, L., Huse, S., Sogin, M. L. & Relman, D. A. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol.* **6**, e280 (2008).
- Walker, A. Say hello to our little friends. *Nature Rev. Microbiol.* **5**, 572–573 (2007).
- Kroghelt, K. A. Bacterial adhesion: genetics, biogenesis, and role in pathogenesis of fimbrial adhesins of *Escherichia coli*. *Rev. Infect. Dis.* **13**, 721–735 (1991).
- Salonen, A. *et al.* Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *J. Microbiol. Methods* **81**, 127–134 (2010).
- Rajilic-Stojanovic, M. *et al.* Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ. Microbiol.* **11**, 1736–1751 (2009).
- Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
- Vanhoutte, T., Huys, G., Brandt, E. d. & Swings, J. Temporal stability analysis of the microbiota in human feces by denaturing gradient gel electrophoresis using universal and group-specific 16S rRNA gene primers. *FEMS Microbiol. Ecol.* **48**, 437–446 (2004).
- Tannock, G. W. *et al.* Analysis of the fecal microflora of human subjects consuming a probiotic product containing *Lactobacillus rhamnosus* DR20. *Appl. Environ. Microbiol.* **66**, 2578–2588 (2000).
- Seksis, P. *et al.* Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut* **52**, 237–242 (2003).
- Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694–1697 (2009).
- Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J. Biol. Chem.* **284**, 24673–24677 (2009).
- Wright, D. P., Rosendale, D. I. & Robertson, A. M. *Prevotella* enzymes involved in mucin oligosaccharide degradation and evidence for a small operon of genes expressed during growth on mucin. *FEMS Microbiol. Lett.* **190**, 73–79 (2000).
- Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int. J. Syst. Evol. Microbiol.* **54**, 1469–1476 (2004).
- Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
- Schwiertz, A. *et al.* Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* **18**, 190–195 (2009).

31. Woodmansey, E. J. Intestinal bacteria and ageing. *J. Appl. Microbiol.* **102**, 1178–1186 (2007).
32. Kovackova, G. & Skorupski, K. The alternative sigma factor σ^F plays an important role in intestinal survival and virulence in *Vibrio cholerae*. *Infect. Immun.* **70**, 5355–5362 (2002).
33. Fujihashi, K. & Kiyono, H. Mucosal immunosenescence: new developments and vaccines to control infectious diseases. *Trends Immunol.* **30**, 334–343 (2009).
34. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
35. Raes, J., Korb, J. O., Lercher, M. J., von Mering, C. & Bork, P. Prediction of effective genome size in metagenomic samples. *Genome Biol.* **8**, R10 (2007).
36. Gibson, G. R. *et al.* Alternative pathways for hydrogen disposal during fermentation in the human colon. *Gut* **31**, 679–683 (1990).
37. Godon, J. J., Zumstein, E., Dabert, P., Habouzit, F. & Moletta, R. Molecular microbial diversity of an anaerobic digester as determined by small-subunit rDNA sequence analysis. *Appl. Environ. Microbiol.* **63**, 2802–2813 (1997).
38. Arumugam, M., Harrington, E. D., Foerster, K. U., Raes, J. & Bork, P. Smash Community: a metagenomic annotation and analysis tool. *Bioinformatics* **26**, 2977–2978 (2010).
39. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
40. Gianoulis, T. A. *et al.* Quantifying environmental adaptation of metabolic pathways in metagenomics. *Proc. Natl Acad. Sci. USA* **106**, 1374–1379 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors are grateful to C. Creevey, G. Falony and members of the Bork group at EMBL for discussions and assistance. We thank the EMBL IT core facility and Y. Yuan for managing the high-performance computing resources. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007–2013): MetaHIT, grant agreement HEALTH-F4-2007-201052, EMBL, the Lundbeck Foundation Centre for Applied Medical Genomics in Personalized Disease Prediction, Prevention and Care (LuCAMP), Novo Nordisk Foundation and the International Science and Technology Cooperation Project in China (0806). Obese/non-obese volunteers for the MicroObes study were recruited from the SU.VI.MAX cohort study coordinated by P. Galan and S. Herberg, and metagenome sequencing was funded by Agence Nationale de la Recherche (ANR); volunteers for MicroAge study were recruited from the CROWNALIFE cohort study coordinated by S. Silvi and A. Cresci, and metagenome sequencing was funded by GenoScope. Ciberehd is funded by the Instituto de Salud Carlos III (Spain). J.R. is supported by the Institute for the encouragement of Scientific Research and Innovation of Brussels (ISIRI) and the Odysseus programme of the Fund for Scientific Research Flanders (FWO). We are thankful to the Human Microbiome Project for generating the reference genomes from human gut microbes and the International Human Microbiome Consortium for discussions and exchange of data.

Author Contributions All authors are members of the Metagenomics of the Human Intestinal Tract (MetaHIT) Consortium. Jun W., F.G., O.P., W.M.d.V., S.B., J.D., Jean W.,

S.D.E. and P.B. managed the project. N.B., F.C., T.H., C.M. and T. N. performed clinical analyses. M.L. and F.L. performed DNA extraction. E.P., D.L.P., T.B., J.P. and E.U. performed DNA sequencing. M.A., J.R., S.D.E. and P.B. designed the analyses. M.A., J.R., T.Y., D.R.M., G.R.F., J.T., J.-M.B., M.B., L.F., L.G., M.K., H.B.N., N.P., J.Q., T.S.-P., S.T., D.T., E.G.Z., S.D.E. and P.B. performed the analyses. M.A., J.R., P.B. and S.D.E. wrote the manuscript. M.H., T.H., K.K. and the MetaHIT Consortium members contributed to the design and execution of the study.

Author Information Raw Sanger read data from the European faecal metagenomes have been deposited in the NCBI Trace Archive with the following project identifiers: MH6 (33049), MH13 (33053), MH12 (33055), MH30 (33057), CD1 (33059), CD2 (33061), UC4 (33113), UC6 (33063), NO1 (33305), NO3 (33307), NO4 (33309), NO8 (33311), OB2 (33313), OB1 (38231), OB6 (38233), OB8 (45929), A (63073), B (63075), C (63077), D (63079), E (63081), G (63083). Contigs, genes and annotations are available to download from http://www.bork.embl.de/Docu/Arumugam_et_al_2011/. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.B. (bork@embl.de) or S.D.E. (dusko.ehrlich@jouy.inra.fr).

MetaHIT Consortium (additional members)

María Antolín¹, François Artiguenave², Hervé M. Blottiere³, Mathieu Almeida³, Christian Brechot^{1,2}, Carlos Cara⁴, Christian Chervaux⁵, Antonella Cultrone³, Christine Delorme³, Gérard Denariáz⁵, Rozenn Dervyn³, Konrad U. Foerster^{6,7}, Carsten Friss⁸, Maarten van de Guchte³, Eric Guedon³, Florence Haimet³, Wolfgang Huber⁶, Johan van Hylckama-Vlieg⁵, Alexandre Jamet³, Catherine Juste³, Ghali Kaci³, Jan Knol⁵, Omar Lakhdari³, Severine Layec³, Karine Le Roux³, Emmanuelle Maguin³, Alexandre Mérieux¹², Raquel Melo Minardi², Christine M'rini¹², Jean Muller⁹, Raish Oozeer⁵, Julian Parkhill¹⁰, Pierre Renault³, Maria Rescigno¹¹, Nicolas Sanchez², Shinichi Sunagawa⁶, Antonio Torrejon¹, Keith Turner¹⁰, Gaetana Vandemeulebrouck³, Encarna Varela¹, Yohanan Winogradsky³ & Georg Zeller⁶

¹Digestive System Research Unit, University Hospital Vall d'Hebron, Ciberehd, 08035 Barcelona, Spain. ²Commissariat à l'Energie Atomique, Genoscope, 91000 Evry, France. ³Institut National de la Recherche Agronomique, 78350 Jouy en Josas, France. ⁴UCB Pharma SA, 28046 Madrid, Spain. ⁵Danone Research, 91120 Palaiseau, France. ⁶European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. ⁷Heidelberger Strasse 24, 64285 Darmstadt, Germany. ⁸Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark. ⁹Institute of Genetics and Molecular and Cellular Biology, CNRS, INSERM, University of Strasbourg, 67404 Illkirch, France. ¹⁰The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. ¹¹Istituto Europeo di Oncologia, 20100 Milan, Italy. ¹²Institut Mérieux, 17 rue Burgelat, 69002 Lyon, France.

Sequential interactions with Sec23 control the direction of vesicle traffic

Christopher Lord^{1*}, Deepali Bhandari^{1*}, Shekar Menon¹, Majid Ghassemian², Deborah Nycz³, Jesse Hay³, Pradipta Ghosh⁴ & Susan Ferro-Novick¹

How the directionality of vesicle traffic is achieved remains an important unanswered question in cell biology. The Sec23p/Sec24p coat complex sorts the fusion machinery (SNAREs) into vesicles as they bud from the endoplasmic reticulum (ER). Vesicle tethering to the Golgi begins when the tethering factor TRAPPI binds to Sec23p. Where the coat is released and how this event relates to membrane fusion is unknown. Here we use a yeast transport assay to demonstrate that an ER-derived vesicle retains its coat until it reaches the Golgi. A Golgi-associated kinase, Hrr25p (CK1 δ orthologue), then phosphorylates the Sec23p/Sec24p complex. Coat phosphorylation and dephosphorylation are needed for vesicle fusion and budding, respectively. Additionally, we show that Sec23p interacts in a sequential manner with different binding partners, including TRAPPI and Hrr25p, to ensure the directionality of ER–Golgi traffic and prevent the back-fusion of a COPII vesicle with the ER. These events are conserved in mammalian cells.

Membrane fusion is mediated by a highly conserved family of membrane proteins called SNAREs. The pairing of a SNARE on the vesicle (v-SNARE) with its cognate SNARE on the target membrane (t-SNARE) is required for fusion¹; however, the same v-SNARE (Sec22p) can act in both anterograde ER–Golgi and retrograde Golgi–ER traffic². This observation implies that factors other than the SNAREs define the direction of membrane flow.

Motifs in the ER–Golgi SNAREs needed for fusion are masked by Sec24p (subunit of the COPII coat) as these fusogens are sorted into ER-derived vesicles³. The COPII coat is assembled on the ER when the activated form of the GTPase Sar1p (Sar1p–GTP) recruits the Sec23p/Sec24p complex by binding to Sec23p, the GTPase-activating protein (GAP) for Sar1p⁴. Polymerization of the coat requires the recruitment of the Sec13p/Sec31p complex (coat outer shell) by the Sec23p/Sec24p complex, which leads to the hydrolysis of GTP on Sar1p and vesicle fission⁴.

The initial interaction of a vesicle with its target membrane is mediated by a class of proteins called tethers that work in conjunction with GTPases of the Rab family⁵. The tethering factor TRAPPI is a multimeric guanine nucleotide exchange factor (GEF) that recruits and activates the Rab GTPase Ypt1p⁶. Previous findings showed that the interaction of TRAPPI with the coat adaptor protein Sec23p is required for vesicle tethering⁶. These studies, however, did not address the question of whether the coat is shed before or after the vesicles bind to the Golgi. Here we show that a Golgi-localized kinase, Hrr25p, displaces purified TRAPPI that is pre-bound to Sec23p and phosphorylates the Sec23p/Sec24p complex. Our findings show that the COPII coat subunit Sec23p interacts in a hierarchical manner with Sar1p, TRAPPI and Hrr25p to ensure the directionality of anterograde membrane flow.

The Golgi inhibits TRAPPI vesicle binding

After COPII vesicles bud from the ER, Sar1p is released from vesicles when GTP is hydrolysed, but the inner and outer shells of the coat are largely retained (Supplementary Fig. 1)⁷. To define the events that

occur after TRAPPI binds to Sec23p, we immobilized TRAPPI on beads and asked when COPII vesicles lose their ability to bind. For these studies, the binding of pro- α -factor-containing vesicles formed with cytosol was considered to be 100% (Fig. 1a). We observed that vesicles formed in the presence of Golgi lost their ability to bind TRAPPI (Fig. 1a). Because the binding of vesicles to TRAPPI is mediated by the COPII coat, this experiment indicates that COPII vesicles lose their ability to bind TRAPPI because the Golgi contains a factor that either releases or modifies the coat. To determine whether vesicles must tether to the Golgi to lose their ability to bind to TRAPPI, we formed vesicles with *bet3-1* mutant fractions. The *bet3-1* mutant, which harbours a mutation in the Bet3p subunit of TRAPPI, is defective in vesicle tethering⁶. The defect in this mutant is partially complemented *in vitro* by the addition of purified recombinant TRAPPI (Supplementary Fig. 2a). Vesicles formed from *bet3-1* donor cells and cytosol, with or without Golgi, bound equally well to the TRAPPI-containing beads (Fig. 1a). These findings indicate that the vesicles must tether to the Golgi to lose their ability to bind to TRAPPI.

COPII vesicle tethering requires TRAPPI, Ypt1p and Uso1p^{5,6}. To determine when COPII vesicles lose their ability to bind to TRAPPI, we blocked vesicle tethering and fusion at several different steps *in vitro* in the presence of Golgi membranes. The pro- α -factor-containing membranes, formed during these blocks, were then tested for their ability to bind to TRAPPI. The transport incompetent vesicles that formed, when Ypt1p function was blocked with anti-Ypt1p antibody, bound efficiently to TRAPPI (Fig. 1b). Disrupting Ypt1p function should also block the recruitment to vesicles of the long coiled-coil tether Uso1p (yeast orthologue of p115), a Ypt1p effector that links donor and acceptor membranes to each other *in vitro*⁵. When we formed vesicles with fractions from the *uso1-1* mutant, transport incompetent vesicles retained their ability to bind TRAPPI at 27 °C and 17 °C (Fig. 1c).

We also blocked fusion *in vitro* with antibody directed against Sec22p (SNARE) or Sly1p, a Sec1-like protein that binds to SNAREs⁸. *In vitro*, neutralizing antibody to Sly1p blocks *trans*-SNARE complex formation⁸. Binding to TRAPPI decreased when vesicles were formed in the

¹Department of Cellular and Molecular Medicine, Howard Hughes Medical Institute, University of California at San Diego, La Jolla, California 92093-0668, USA. ²Department of Chemistry and Biochemistry, Biomolecular and Proteomics Mass Spectrometry Facility, University of California at San Diego, La Jolla, California 92093, USA. ³Division of Biological Sciences, The University of Montana, Missoula, Montana 59812, USA. ⁴Department of Medicine, University of California at San Diego, La Jolla, California, 92093-0651, USA.

*These authors contributed equally to this work.

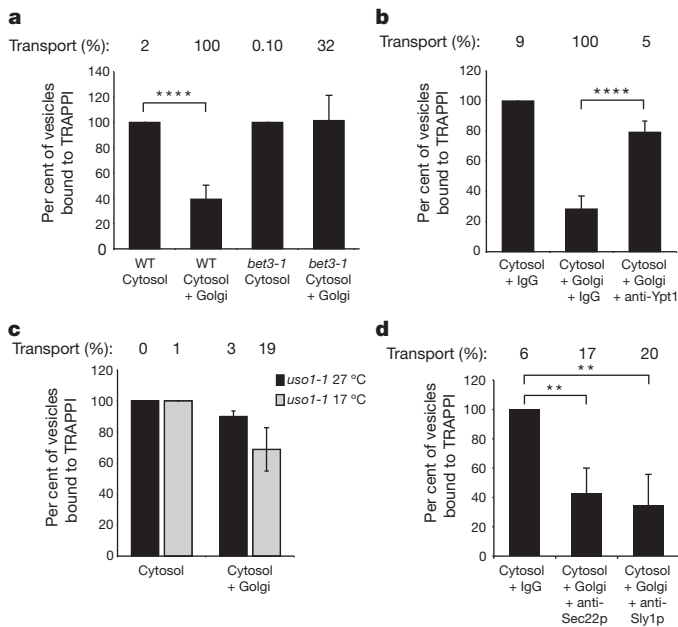


Figure 1 | COPII vesicles lose their ability to bind TRAPPI after Uso1p acts. **a**, Vesicles formed *in vitro* with cytosol with or without Golgi were incubated with TRAPPI-containing beads and the precipitated radiolabelled cargo was counted. Cytosol and Golgi were isolated from either wild type (WT) or the *bet3-1* mutant. Error bars represent standard deviation (s.d.), $N = 6$. In **a–d**, the per cent transport observed for the total reaction is reported above the bar graphs. **b**, **d**, Cytosol and Golgi fractions, derived from wild type, were formed with: anti-Ypt1p (12 μ g), anti-Sec22p (12 μ g), anti-Sly1p (20 μ g) antibodies, or IgG (12 μ g or 20 μ g, respectively). Error bars represent s.d., $N = 4$. **c**, Vesicles were formed with *uso1-1* fractions at 27 °C or 17 °C and incubated with TRAPPI-containing beads. $N = 2$, bars show the range. ** $P < 0.01$, **** $P < 0.0001$ Student's *t*-test.

presence of anti-Sec22p or anti-Sly1p antibodies (Fig. 1d). These findings indicate that COPII vesicles lose their ability to bind TRAPPI after Uso1p functions, but before *trans*-SNARE complex formation. Uso1p does not seem to have a role in vesicle uncoating, as the membrane and soluble pools of Sec23p were unaltered in the *uso1-1* mutant *in vivo* (Supplementary Fig. 3a) and Uso1p/p115 did not release Sec23 from membranes *in vitro* (see Supplementary Figs 3b, c). Together, these findings indicate that COPII vesicles retain their coat until they reach the Golgi.

Hrr25p phosphoregulates Sec23p/Sec24p

Despite the fact that only $36 \pm 0.65\%$ of wild-type vesicles uncoat *in vitro* in the presence of Golgi membranes (Supplementary Fig. 3d), $61 \pm 4\%$ of the vesicles lose their ability to bind to TRAPPI (Fig. 1a). This observation indicates that the inability of COPII vesicles to bind TRAPPI is not just a consequence of vesicle uncoating. Because the COPII coat inner shell is known to be phosphorylated in mammalian cells⁹, we considered the possibility that phosphorylation of Sec23p may block the ability of COPII vesicles to bind TRAPPI.

To identify a kinase that could phosphorylate Sec23p, we searched the yeast database (<http://www.yeastgenome.org>) for an essential (Supplementary Fig. 3e) Golgi-localized kinase. Only one kinase, Hrr25p, was found to have an orthologue (CKI δ , human orthologue of casein kinase I δ) that localizes at the Golgi and ER–Golgi interface in mammalian interphase cells¹⁰. Inhibiting CKI δ function was reported to block ER–Golgi traffic¹¹, and a mutation that reduced the kinase activity of Hrr25p was shown to suppress the temperature-sensitive COPII vesicle budding defect in the *sec12-4* mutant¹². Although these results implicated Hrr25p/CKI δ in ER–Golgi traffic, its role in membrane traffic is not well defined.

Hrr25p is known to reside in the nucleus in G1-arrested cells and, like CKI δ , it is found at the spindle pole body (SPB) in nocodazole-treated (M-phase) cells^{10,13}. When we visualized Hrr25p–RFP (genomic copy tagged with red fluorescent protein) in asynchronously grown cells, however, the majority (>95%) was found on punctate structures that largely colocalize with early (Vrg4p) and late Golgi (Sec7p) markers (Fig. 2a). Occasionally, we observed Hrr25p–RFP in the nucleus, at puncta along the nuclear envelope (presumably the SPB), and at the bud neck and cortex, as previously reported^{13,14}. Consistent with its Golgi localization, all of the haemagglutinin (HA)-tagged Hrr25p co-fractionated with membranes in differential fractionation experiments (Fig. 5a, bottom).

Both glutathione S-transferase (GST)-tagged Sec23p and GST–Sec24p were phosphorylated by His₆-tagged Hrr25p (Fig. 2b and Supplementary Fig. 3f, compare lanes 1 and 2), but not catalytically inactive His₆–Hrr25p(K38A) *in vitro* (Fig. 2b and Supplementary Fig. 3f, lane 3). Phosphorylation *in vivo* was examined with a conditional allele of *hrr25* (ref. 14). In this mutant, HA-tagged HRR25^{degron} was placed behind the inducible GAL promoter as the sole copy of the gene. When this strain is grown in galactose, HA–Hrr25p^{degron} is expressed. In the presence of glucose, however, the expression of HA–Hrr25p^{degron} ceases and the protein is rapidly degraded. After 10 h in glucose, when a delay in trafficking of carboxypeptidase Y between the ER–Golgi was observed, most of the Hrr25p^{degron} was degraded (not shown). Lysates prepared from the *hrr25* mutant grown with galactose (+Hrr25p) or glucose (–Hrr25p) (Supplementary Fig. 3g) were precipitated with anti-Sec24p (Fig. 2c, lanes 1 and 3), or IgG (Fig. 2c, lanes 2 and 4). Western blot analysis of the immunoprecipitates with anti-phospho-Ser/Thr antibody revealed that phospho-Sec23p and phospho-Sec24p could only be detected *in vivo* when Hrr25p was expressed (Fig. 2c, lane 1).

As Sec23p binds to both TRAPPI and Hrr25p, we determined whether TRAPPI and Hrr25p compete for binding to Sec23p. To do this, the six subunits of the TRAPPI complex were co-expressed in bacteria and purified from bacterial lysates as described previously⁶. When similar amounts of purified His₆–Hrr25p and TRAPPI were incubated together, Hrr25p effectively competed with TRAPPI for binding to purified GST–Sec23p (Supplementary Fig. 4a, compare lanes 3–5 to GST controls in lanes 1 and 2). Binding was not dependent on kinase activity, as His₆–Hrr25p(K38A) bound as efficiently as His₆–Hrr25p (not shown). Because His₆–Hrr25p binds with higher affinity to GST–Sec23p ($K_d = 0.043 \pm 0.009$ μ M; Supplementary Fig. 4b) than TRAPPI ($K_d = 0.63 \pm 0.15$ μ M; Supplementary Fig. 4c), we determined whether it displaces TRAPPI that is pre-bound to GST–Sec23p. Increasing amounts of His₆–Hrr25p were mixed with GST–Sec23p beads pre-incubated with saturating amounts of TRAPPI (Fig. 2d). When the concentration of Hrr25p was increased (Fig. 2d, lanes 2–5), TRAPPI was released from the beads (Fig. 2d, top, lanes 2–5) into the supernatant (Fig. 2d, bottom, lanes 2–5) as His₆–Hrr25p bound to GST–Sec23p (Fig. 2d, middle). These findings show that Hrr25p and TRAPPI compete for binding to Sec23p, and indicate that Hrr25p could displace TRAPPI from Sec23p when COPII vesicles tether to the Golgi. Consistent with the possibility that phosphorylation of the coat blocks the binding of TRAPPI to Sec23p, we observed a decrease in the binding of TRAPPI to GST–Sec23p that contained phosphomimetic mutations at two phosphorylation sites (see later and Supplementary Fig. 5b).

Conservation of phosphorylation sites in Sec23p

Because Hrr25p phosphorylates Sec23p more efficiently than Sec24p, we focused on Sec23p for subsequent studies. Three Hrr25p phosphorylation sites were identified in Sec23p by mass spectrometry: T555, S742 and T747. Two of these phosphorylated residues, S742 and T747, are conserved from yeast to man and were analysed further (Fig. 3a, top). The T747 residue is a known Sar1p contact site, whereas S742 is within a disordered loop in the established structure of Sec23p

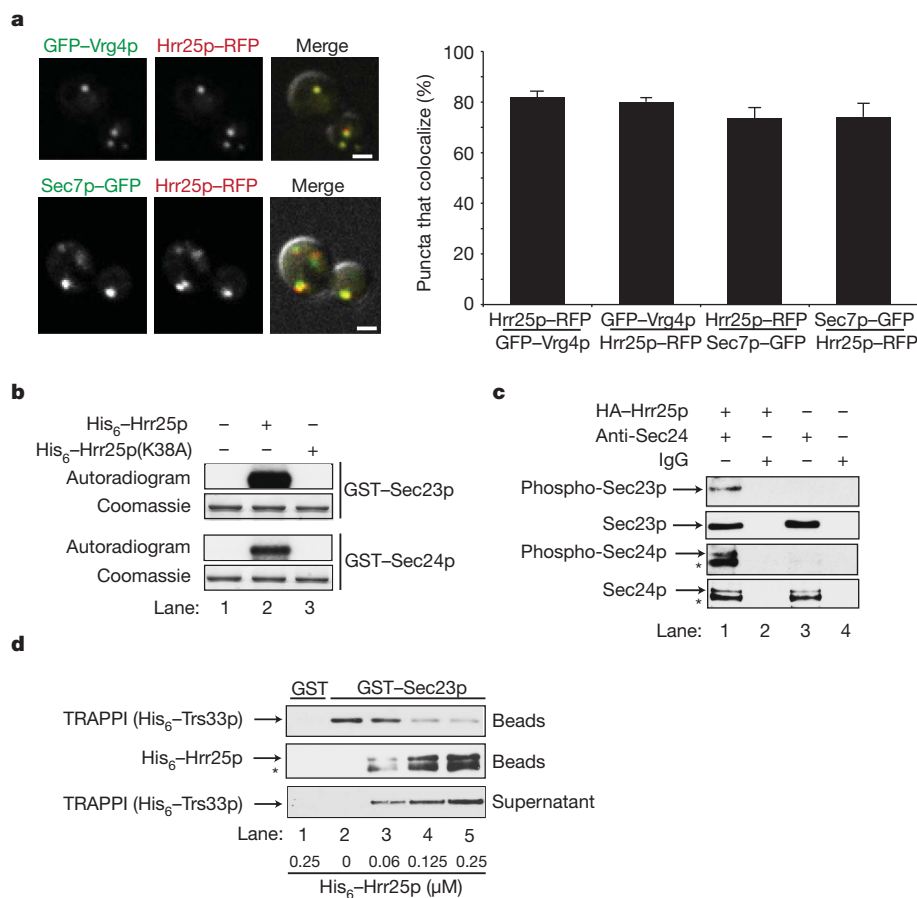


Figure 2 | Hrr25p resides on the Golgi and phosphorylates Sec23p/Sec24p. **a**, Hrr25p-RFP colocalizes with GFP-Vrg4p (top) and Sec7p-GFP (bottom). The green and red channels are merged with the differential interference contrast (DIC) image (right panel). The puncta that colocalize (s.d., $N = 3$ experiments) are shown on the right. Scale bar, 2 μm. **b**, GST-Sec23p and GST-Sec24p were incubated without (lane 1), or with His₆-Hrr25p (lane 2) or His₆-Hrr25p(K38A) (lane 3) and γ P³²-ATP. The autoradiogram and coomassie-stained gel are shown. **c**, Lysates expressing (lanes 1 and 2) or not expressing HA-Hrr25p (lanes 3 and 4) were immunoprecipitated with

anti-Sec24 antibody (lanes 1 and 3) or IgG (lanes 2 and 4) and immunoblotted with anti-phospho-Ser/Thr, anti-Sec23p and anti-Sec24p antibodies. **d**, TRAPPI, pre-bound to GST-Sec23p-containing beads (top), was incubated with increasing concentrations of His₆-Hrr25p. The beads were pelleted and the amount of TRAPPI in the supernatant (bottom) and pellet (top) was assessed. The Hrr25p that bound to the beads (middle) was also measured. TRAPPI, or Hrr25p, did not bind to GST (lane 1). The starred bands in **c** and **d** are degradation products of Sec24p and His₆-Hrr25p.

complexed with Δ23-Sar1p-GTP (PDB accession 2QTV)¹⁵. Initially, we used computational modelling to predict the consequences of phosphorylation at these sites. We added phosphates on S742 and T747 of Sec23p using Molsoft-ICM software and modified their orientations by several rounds of Monte Carlo optimization. Phosphorylation at T747 presented significant steric clashes with the surface of Sar1p in all possible orientations and was deemed incompatible with Sec23p binding to Sar1p-GTP (Fig. 3a, bottom). Phosphorylated S742 is also located at the Sec23p-Sar1p interface, but its location within a flexible loop limited our ability to predict consequences (Fig. 3a, bottom)¹⁵. We found that His₆-Δ23-Sar1p-GTPγS bound preferentially to GST-Sec23p (Supplementary Fig. 5a), and failed to bind GST-Sec23p harbouring the phosphomimetic S742D, T747E or S742D/T747E (ST-DE) mutations (Fig. 3b, top, lanes 1–4). Binding of His₆-Sec24p to GST-Sec23p was unaffected by the phosphomimetic mutations (Fig. 3b, bottom).

If Hrr25p phosphorylates Sec23p at Sar1p contact sites, the addition of His₆-Hrr25p to the transport assay should disrupt the binding of Sec23p to Sar1p-GTP and inhibit vesicle budding *in vitro*. When His₆-Hrr25p was added to the assay at the beginning of the reaction, budding was inhibited as the concentration of His₆-Hrr25p was increased (Fig. 3c). Inhibition was dependent on kinase activity, as no effect was seen with catalytically inactive His₆-Hrr25p(K38A)

(Fig. 3c). Consistent with this observation, a yeast strain harbouring the S742D/T747E mutations disrupted vesicle budding and fusion *in vitro* (Fig. 3d) and displayed a severe growth defect at 33 °C (Supplementary Fig. 5c). The defect in fusion may be the consequence of blocking the cycling of Sec23p on and off membranes (see Fig. 5a).

Because Hrr25p phosphorylates Sec23p at Sar1p contact sites, and Hrr25p competes with TRAPPI for binding to Sec23p, we wanted to address whether TRAPPI also competes with His₆-Δ23-Sar1p-GTPγS for binding to Sec23p. When we incubated a constant amount of His₆-Δ23-Sar1p-GTPγS with increasing amounts of TRAPPI, TRAPPI effectively competed with Δ23-Sar1p-GTPγS for binding to Sec23p (Fig. 4a, top). Similar results were obtained when increasing amounts of His₆-Δ23-Sar1p-GTPγS were incubated with a constant amount of TRAPPI (Fig. 4a, bottom). These results imply that TRAPPI and Sar1p-GTP bind to the same or overlapping site(s) on Sec23p. Consistent with this hypothesis, we found a decrease in the binding of TRAPPI to GST-Sec23p harbouring the S742D/T747E mutations (Supplementary Fig. 5b). Together, these findings imply that TRAPPI binds to Sec23p after Sar1p-GTP is released from membranes. This event seems to be conserved in higher eukaryotes (Fig. 4b), as we could only detect mBet3 on immuno-isolated tagged mammalian COPII vesicles (VSV-G-Myc) formed *in vitro* with GTP (–Sar1), but not the non-hydrolysable GTP analogue GMP-PNP (+Sar1) or control vesicles that lacked VSV-G-Myc.

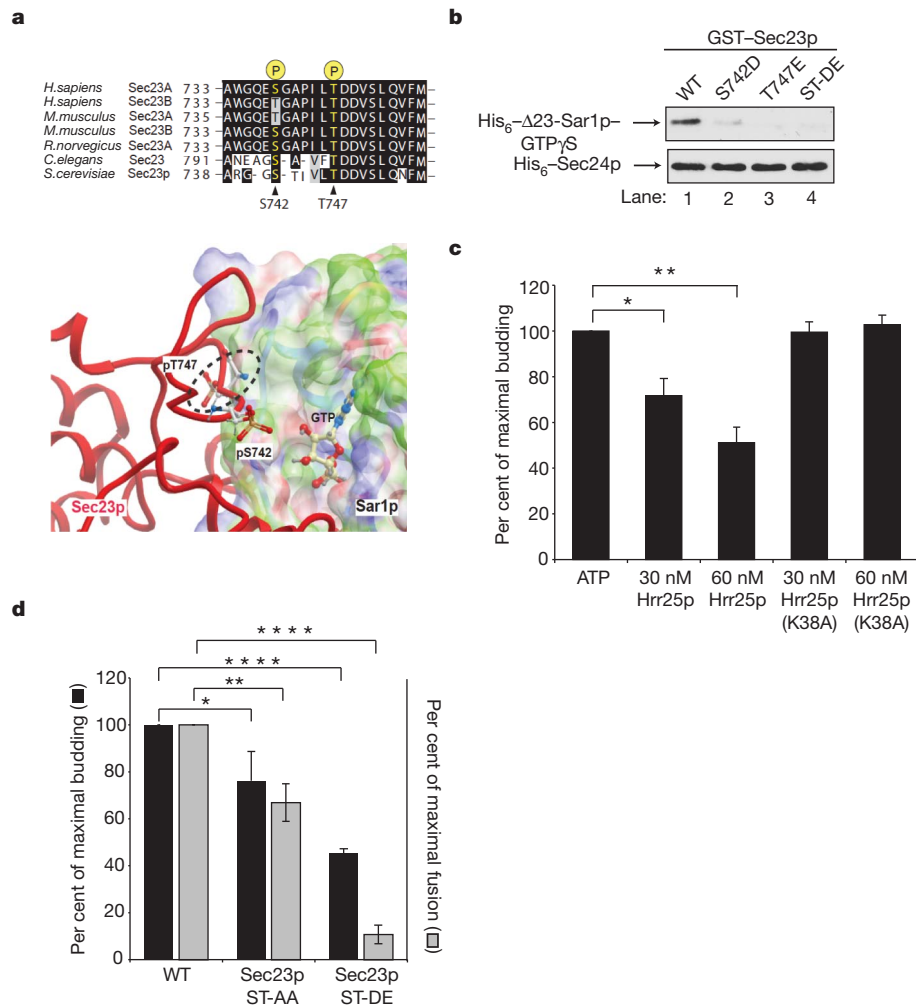


Figure 3 | Phosphorylation of S742 and T747 blocks ER-Golgi traffic *in vitro*. **a**, Top, the sequence flanking S742 and T747 in Sec23p was aligned with Sec23 orthologues. Bottom, phosphorylated S742 and T747 in Sec23p (red) are located at its interface with Sar1p. The electrostatic potential of the Sec23p-binding surface of Sar1p is coloured according to solvation properties of the residues (white, hydrophobic; green, polar; blue, basic; red, acidic). The dotted ellipsoid marks the steric clash between Sec23p and Sar1p-GTP. **b**, Top, wild-type (WT) and mutant GST-Sec23p fusion proteins were incubated with 10 nM of His₆-Δ23-Sar1p-GTPγS or His₆-Sec24p. A truncated form of Sar1p

was used for these studies because the full-length protein aggregates¹⁵.

c, Increasing amounts of purified His₆-Hrr25p or His₆-Hrr25p(K38A) were added *in vitro* at the beginning of a complete transport reaction and vesicle budding was measured. Error bars represent s.d., *N* = 4. **d**, Fractions prepared from wild type and strains containing non-phosphorylatable (ST-AA) and phosphomimetic (ST-DE) Sec23p were assayed for vesicle budding and fusion as before⁶. Error bars represent standard error of the mean (s.e.m.), *N* = 4.

P* < 0.05, *P* < 0.01, *****P* < 0.0001 Student's *t*-test.

Loss of Hrr25p activity inhibits fusion

To address the role of Hrr25p phosphorylation in ER-Golgi traffic *in vitro*, we used the ATP competitive inhibitor IC261, which selectively inhibits the highly conserved kinase domain of CK1δ^{11,16}. As seen in Fig. 4c, increasing concentrations of IC261 inhibited fusion. To address whether IC261 inhibits membrane fusion or vesicle tethering, we formed vesicles in the presence of Golgi with or without inhibitor, and then fractionated the reaction product on a sucrose velocity gradient that separates free vesicles (Supplementary Fig. 2b, top) from vesicles that are bound to the Golgi (Supplementary Fig. 2b, bottom). Subsequently, each fraction was treated with ConA Sepharose to precipitate radiolabelled pro-α-factor. Most of the vesicles bound to the Golgi in the presence of IC261, indicating that the inhibitor largely blocks fusion and not tethering (Supplementary Fig. 2b, bottom).

Interestingly, when the transport assay was performed with concentrations of IC261 that inhibited fusion, a stimulation in vesicle budding was observed (Fig. 4d). This finding indicates that dephosphorylation is needed for budding and is consistent with an earlier report showing that phosphorylated mammalian Sec24 cannot form a pre-budding complex⁹. A prediction of this result is that loss of kinase

activity should stimulate cargo export *in vivo* (see later) and could explain why a kinase loss-of-function mutation was identified as a suppressor of the *sec12-4* mutant¹². We were unable to address the role of CK1δ *in vitro* because the mammalian COPII vesicle tethering assay did not work robustly at the lower ATP concentrations needed to test the inhibitor.

To address whether phosphorylation and dephosphorylation alter the distribution of Sec23p on membranes *in vivo*, a differential fractionation experiment was performed with the conditional *hrr25* mutant after growth in galactose- or glucose-containing medium. The SNARE Bos1p served as a membrane marker for these studies. This analysis revealed the presence of a soluble pool of Sec23p when HA-Hrr25p was expressed (Fig. 5a, lanes 1–3). In the absence of HA-Hrr25p, however, all of the Sec23p was membrane-bound (Fig. 5a, lanes 4–6). Although Hrr25p activity seems to have a role in releasing Sec23p from membranes *in vivo*, we found it was not sufficient to uncoat the vesicles in the absence of Golgi membranes *in vitro* (data not shown).

In mammalian cells, COPII vesicles fuse to each other or with COPI (Golgi) vesicles to form a pre-Golgi compartment that matures into a

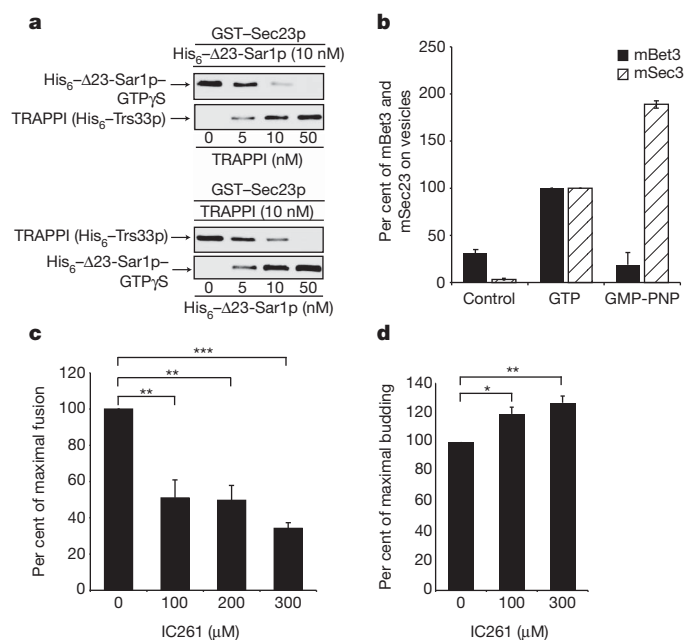


Figure 4 | TRAPPI and Sar1p-GTP cannot bind to Sec23p simultaneously. **a**, GST-Sec23p beads were incubated with 10 nM of His₆-Δ23Sar1p-GTPγS and increasing concentrations of TRAPPI (top), or 10 nM TRAPPI and increasing concentrations of His₆-Δ23Sar1p-GTPγS (bottom). **b**, mBet3 is on COPII vesicles formed with GTP (–Sar1), but not GMP-PNP (+Sar1). Vesicles formed *in vitro* with GTP or GMP-PNP were immunoprecipitated with an antibody against the cargo VSV-G-Myc and blotted for mBet3 and mSec23. The data was normalized to cargo yield (see Methods). *N* = 2, bars show the range. **c**, *In vitro* transport was performed with increasing concentrations of IC261 and fusion was measured. Error bars represent s.d., *N* = 4. **d**, The *in vitro* transport assay was performed as in Fig. 4c with IC261 and budding was measured. Error bars represent s.d., *N* = 3. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 Student's *t*-test.

Golgi¹⁷. To address the role of CK1δ *in vivo*, we accumulated tsO45 VSV-G-GFP in the ER of NRK cells at 40 °C and then shifted the cells in the presence and absence of IC261 to 15 °C, a temperature that selectively slows traffic at the pre-Golgi compartment step¹⁸. In the presence of inhibitor, the pre-Golgi (marked by rbet1) but not early Golgi (marked by gpp130) was markedly dispersed (Fig. 5b). Consistent with the observation that inhibiting CK1δ function stimulates COPII vesicle budding and blocks fusion, VSV-G-GFP was more rapidly depleted from the ER and concentrated at peripheral sites of the pre-Golgi compartment (Fig. 5b, c), the site where COPII vesicles fuse. The VSV-G-GFP remained at the peripheral sites in the IC261-treated cells and failed to concentrate in the peri-Golgi region (Fig. 5d, e). Together, these findings imply that the events we describe here are conserved in higher cells.

Discussion

The CKI family of kinases represents a unique group of highly conserved serine/threonine kinases that regulate a variety of cellular processes, including membrane traffic^{11,12}. Here we report that Sec23p, a component of the inner shell of the COPII coat, sequentially interacts with three different binding partners, Sar1p, TRAPPI and Hrr25p, to control the direction of ER–Golgi traffic. These interactions define three different stages in vesicle traffic: budding, tethering and a pre-fusion step.

As Sar1p is required for fission^{19,20}, our findings imply that TRAPPI can only bind to Sec23p after vesicle fission and the release of Sar1p from membranes (Fig. 6, (1)). This ensures that COPII vesicle tethering is only initiated after a vesicle buds from the ER. Subsequently, TRAPPI activates Ypt1p on the vesicle (Fig. 6, (2)). Genetic studies and a kinetic analysis of GEF activity have revealed that TRAPPI is

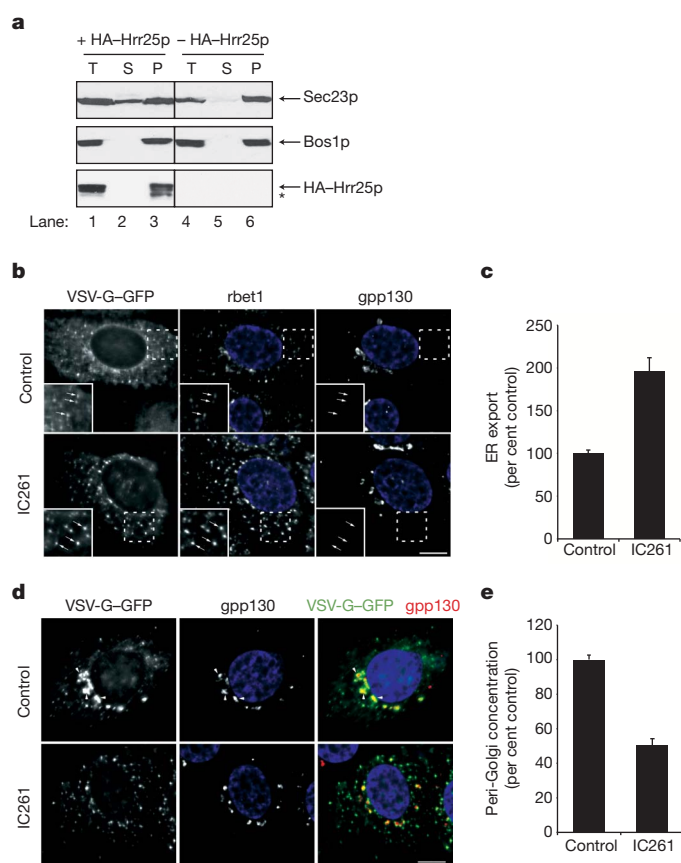


Figure 5 | In the presence of IC261, cargo is exported more rapidly but remains at peripheral pre-Golgi sites. **a**, The *hrr25* mutant was grown in galactose (lanes 1–3), or glucose (lanes 4–6) supplemented media for 10 h. Total (T) lysates were centrifuged at 150,000g and the supernatant (S) and pellet (P) fractions were analysed. The starred band is a degradation product of HA-Hrr25p. **b**, NRK cells that accumulated VSV-G-GFP in the ER at 40 °C were shifted to 15 °C for 30 min in the presence of DMSO (top) or 100 μM IC261 (bottom) to allow cargo transit from the ER to the ER–Golgi interface. Left, VSV-G-GFP fluorescence; middle, rbet1; right, gpp130. Insert in left corner is an expansion of the dotted box. Arrows, peripheral cargo-containing punctate structures that colocalize with rbet1 but not gpp130. **c**, Quantification, from Fig. 5b, of VSV-G-GFP fluorescence in pre-Golgi structures divided by VSV-G-GFP remaining in the ER (see Methods for calculation). Error bars represent s.e.m., *N* = 20 cells per bar, *P* < 0.0001 Student's *t*-test. **d**, Same as **b** only cells were incubated at 15 °C for 60 min. Left, VSV-G-GFP fluorescence; middle, gpp130; right, merge of VSV-G-GFP (green) and gpp130 (red). Arrowheads, cargo that concentrated in the Golgi area. **e**, Quantification, from Fig. 5d, of VSV-G-GFP fluorescence in punctate structures overlapping the Golgi divided by total fluorescence in punctate structures (see Methods for calculation). Error bars represent s.e.m., *N* = 20 cells per bar, *P* < 0.0001 Student's *t*-test. The nuclei are stained with DAPI. Scale bars, 10 μm.

more than a GEF²¹. GEFs typically release Rabs soon after they activate them. TRAPPI, however, forms a relatively stable ternary complex (TRAPPI–Ypt1p–nucleotide) with the Rab²¹, implying that it is also a Ypt1p effector. In parallel, the pool of Ypt1p–GTP that is released from TRAPPI can then recruit the long coiled-coil tether Uso1p (Fig. 6, (3)). When Uso1p bends, the vesicle comes into proximity with the Golgi, triggering the release of TRAPPI from the vesicle (Fig. 6, (3)). Hrr25p, which concentrates on the Golgi in yeast, could facilitate this release. Phosphorylation of the Sec23p/Sec24p complex by Hrr25p may be required, but is not sufficient for COPII vesicle uncoating. Another kinase could also be involved in this event as Sec31p, a known phosphoprotein²², seems to be phosphorylated by a different kinase¹² and Hrr25p cannot uncoat COPII vesicles *in vitro*.

Fusogenic SNARE motifs that bind to the coat must be unmasked before *trans*-SNARE complex formation can proceed at the target

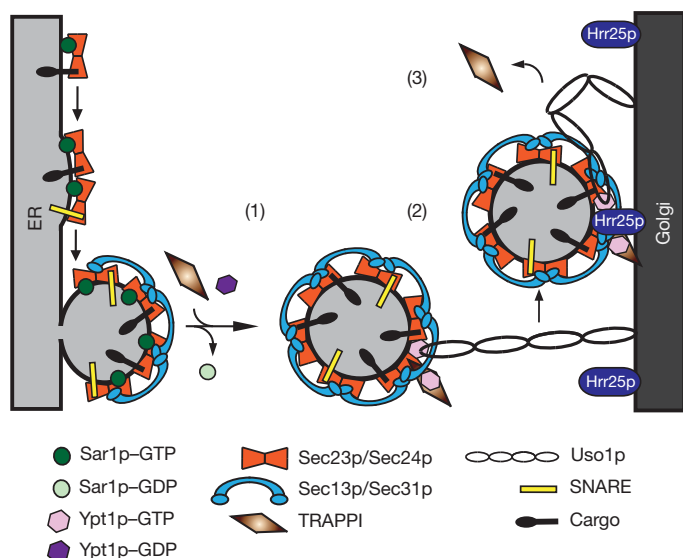


Figure 6 | Sec23p ensures the direction of ER-Golgi traffic. (1) TRAPPI binds to Sec23p after GTP is hydrolysed on Sar1p. (2) Ypt1p is activated by TRAPPI and Uso1p is recruited to the vesicle. Subsequently, Uso1p binds to the Golgi. (3) When Uso1p bends, it brings the vesicle to the Golgi. TRAPPI is then released from the vesicle and Hrr25p phosphorylates the Sec23p/Sec24p complex.

membrane³. Phosphorylation of Sec24p by Hrr25p/CKI δ could have a role in disengaging the SNAREs from the coat at the Golgi. As the COPII coat only acts in anterograde traffic⁴, the compartmentalization of a kinase that regulates membrane fusion at the Golgi, ensures an ER-Golgi v-SNARE will only pair with its cognate t-SNARE. The directionality imposed by this cycle also prevents the back-fusion of a COPII vesicle with the ER. The findings we report here describe a new role for Hrr25p/CKI δ that may extend to other CKI family members and coats.

METHODS SUMMARY

Yeast COPII vesicles were formed *in vitro* with donor cells, cytosol, with or without Golgi for 90 min at 20 °C or 27 °C or 120 min at 17 °C as described previously⁶. The vesicle binding assays were performed as described previously⁶. Additional information is provided in Methods.

Protein and antibody purifications were performed as before⁶. Mass spectrometry analysis, *in vitro* bindings assays, kinase and immunoprecipitation assays, microscopy, the construction of phosphomimetic mutations and all studies with mammalian cells are described in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 September 2010; accepted 23 February 2011.

Published online 1 May 2011.

1. Jahn, R. & Scheller, R. H. SNAREs—engines for membrane fusion. *Nature Rev. Mol. Cell Biol.* **7**, 631–643 (2006).
2. Lewis, M. J., Rayner, J. C. & Pelham, H. R. A novel SNARE complex implicated in vesicle fusion with the endoplasmic reticulum. *EMBO J.* **16**, 3017–3024 (1997).

3. Mossessova, E., Bickford, L. C. & Goldberg, J. SNARE selectivity of the COPII coat. *Cell* **114**, 483–495 (2003).
4. Lee, M. C. S., Miller, E. A., Goldberg, J., Orci, L. & Schekman, R. Bi-directional protein transport between the ER and Golgi. *Annu. Rev. Cell Dev. Biol.* **20**, 87–123 (2004).
5. Whyte, J. R. C. & Munro, S. Vesicle tethering complexes in membrane traffic. *J. Cell Sci.* **115**, 2627–2637 (2002).
6. Cai, H. *et al.* TRAPPI tethers COPII vesicles by binding the coat subunit Sec23. *Nature* **445**, 941–944 (2007).
7. Barlowe, C. *et al.* COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. *Cell* **77**, 895–907 (1994).
8. Flanagan, J. J. & Barlowe, C. Cysteine-disulfide cross-linking to monitor SNARE complex assembly during endoplasmic reticulum-Golgi transport. *J. Biol. Chem.* **281**, 2281–2288 (2006).
9. Dudognon, P., Maeder-Garavaglia, C., Carpentier, J. & Paccaud, J. Regulation of a COPII component by cytosolic O-glycosylation during mitosis. *FEBS Lett.* **561**, 44–50 (2004).
10. Milne, D. M., Looby, P. & Meek, D. W. Catalytic activity of protein kinase CK1 δ (casein kinase 1 δ) is essential for its normal subcellular localization. *Exp. Cell Res.* **263**, 43–54 (2001).
11. Yu, S. & Roth, M. G. Casein kinase I regulates membrane binding by ARF GAP1. *Mol. Biol. Cell* **13**, 2559–2570 (2002).
12. Murakami, A., Kimura, K. & Nakano, A. The inactive form of a yeast casein kinase I suppresses the secretory defect of the sec12 mutant. *J. Biol. Chem.* **274**, 3804–3810 (1999).
13. Lusk, C. P. *et al.* Nup53p is a target of two mitotic kinases, Cdk1p and Hrr25p. *Traffic* **8**, 647–660 (2007).
14. Kafadar, K. A., Zhu, H., Snyder, M. & Cyert, M. S. Negative regulation of calcineurin signaling by Hrr25p, a yeast homolog of casein kinase I. *Genes Dev.* **17**, 2698–2708 (2003).
15. Bi, X., Corpina, R. A. & Goldberg, J. Structure of the Sec23/24-Sar1 pre-budding complex of the COPII vesicle coat. *Nature* **419**, 271–277 (2002).
16. Mashhoon, N. *et al.* Crystal structure of a conformation-selective casein kinase-1 inhibitor. *J. Biol. Chem.* **275**, 20052–20060 (2000).
17. Xu, D. & Hay, J. C. Reconstitution of COPII vesicle fusion to generate a pre-Golgi intermediate compartment. *J. Cell Biol.* **167**, 997–1003 (2004).
18. Schweizer, A. *et al.* Identification of an intermediate compartment involved in protein transport from endoplasmic reticulum to Golgi apparatus. *Eur. J. Cell Biol.* **53**, 185–196 (1990).
19. Bielli, A. *et al.* Regulation of Sar1 NH₂ terminus by GTP binding and hydrolysis promotes membrane deformation to control COPII vesicle fission. *J. Cell Biol.* **171**, 919–924 (2005).
20. Lee, M. C. S. *et al.* Sar1p N-terminal helix initiates membrane curvature and completes the fission of a COPII vesicle. *Cell* **122**, 605–617 (2005).
21. Chin, H. F. *et al.* Kinetic analysis of the guanine nucleotide exchange activity of TRAPP, a multimeric Ypt1p exchange factor. *J. Mol. Biol.* **389**, 275–288 (2009).
22. Salama, N. R., Chuang, J. S. & Schekman, R. W. Sec31 encodes an essential component of the COPII coat required for transport vesicle budding from the endoplasmic reticulum. *Mol. Biol. Cell* **8**, 205–217 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Schekman, M. Cyert, L. Miller, B. Glick, M. Lowe and E. Mizuno-Yamasaki for strains, constructs, purified protein and antibody; K. Reinisch for discussions; S. Chen, H. Cai and M. Garcia-Marcos for advice; W. Zhou and A. Loughheed for technical assistance. This work was supported by the Howard Hughes Medical Institute. Salary support for S.F.-N., D.B. and S.M. was provided by the Howard Hughes Medical Institute, P.G. was funded by the Burroughs Wellcome Fund and M.G. by the SRP Super Fund Research Program. Work at the University of Montana was supported by National Institutes of Health grant GM-059378 (to J.C.H.) and the National Institutes of Health COBRE Center grant RR-015583.

Author Contributions C.L., D.B., S.M., D.N. and J.H. performed experiments and analysed data. M.G. did the mass analysis and P.G. did the computational modelling and analysed data. S.F.-N. directed the project, analysed data and wrote the paper. C.L., D.B., J.H. and P.G. co-wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.F.-N. (sfnovick@ucsd.edu).

METHODS

Yeast *in vitro* transport and vesicle binding assays. For the vesicle binding assays, the permeabilized cells were pelleted after the *in vitro* transport reaction as described previously⁶. The conditions used to remove the cells did not pellet slowly sedimenting membranes that contain radiolabelled cargo. The supernatant was transferred to a new tube that contained 40 µl of a 50% slurry of TRAPPI-containing beads⁶. The final volume of the reaction was adjusted to 500 µl with TBPS (25 mM HEPES (pH 7.2), 115 mM potassium acetate, 2.5 mM magnesium acetate, 250 mM sorbitol plus protease inhibitors) and the reaction was incubated for 2 h at 4 °C. The beads were washed three times with 750 µl of TBPS and counted. The counts from a no vesicle control reaction (reaction with apyrase with or without Golgi) were subtracted as background. The binding of vesicles formed with cytosol was considered to be 100%, and the amount of binding (with or without Golgi) was adjusted to equal Concanavalin A (ConA) counts (equal vesicles). Two-stage transport assays were also performed. Vesicles were formed in the absence of Golgi, the cells were pelleted and an equal number of vesicles were incubated with or without Golgi before the binding assay was performed.

For reactions with IC261, the ATP concentration of the 10× ATP stock was lowered to 1.7 mM ATP. For the vesicle tethering assay, free vesicles were separated from vesicles that bound to the Golgi on a sucrose velocity gradient as described previously⁶.

***In vitro* kinase assay.** Purified GST, GST–Sec23p and GST–Sec24p (5 µg), immobilized on beads, were incubated with 250 ng of His₆–Hrr25p or catalytically inactive His₆–Hrr25p(K38A) or no kinase in kinase assay buffer (50 mM HEPES pH 7.4, 2 mM EDTA, 10 mM MgCl₂, 1 mM DTT, 5 mM cold ATP, 2.5 µCi γP³²-ATP, 100 µM sodium orthovanadate, 10 mM sodium fluoride, 10 mM sodium pyrophosphate and protease inhibitors) for 1 h at 30 °C. The beads were washed twice with 1× PBS and eluted in 25 µl of sample buffer by heating to 100 °C for 5 min. The samples were analysed by autoradiography.

Nucleotide loading of His₆–Δ23–Sar1p. His₆–Δ23–Sar1p was loaded with the desired nucleotide (GDP or GTPγS) overnight at 4 °C in the following buffer: 20 mM HEPES pH 7.2, 150 mM NaCl, 1 mM MgCl₂ with 0.1 mM of nucleotide.

***In vitro* binding assays with recombinant proteins.** For the binding experiments with the phosphomimetic mutations in GST–Sec23p, equimolar amounts (0.1 µM) of GST, GST–Sec23p beads with or without phosphomimetic mutations were incubated with either 5–10 nM of His₆–Δ23–Sar1p–GTPγS (nucleotide was loaded as described earlier) or 10 nM of purified TRAPPI⁶ in binding buffer I (25 mM HEPES pH 7.2, 150 mM NaCl, 2% Triton X-100, 1 mM DTT, 2 mM EDTA, 0.5 mM MgCl₂ and protease inhibitors) for 3–4 h at 4 °C. The beads were washed three times with binding buffer I and eluted in 25 µl of sample buffer by heating to 100 °C for 5 min.

For the TRAPPI displacement assay, 0.5 µM of TRAPPI was incubated overnight at 4 °C in binding buffer I with 0.1 µM of GST–Sec23p or GST. The next day, the beads were washed three times in binding buffer I and resuspended in binding buffer II (25 mM HEPES pH 7.2, 150 mM NaCl, 0.1% Triton X-100, 1 mM DTT, 2 mM EDTA, 0.5 mM MgCl₂ and protease inhibitors). Increasing amounts (0–0.25 µM) of purified His₆–Hrr25p were added to the reactions and incubated for 4 h at 4 °C. Subsequently, the supernatants were aliquoted into fresh tubes and heated in sample buffer to 100 °C for 5 min. The beads were washed three times with binding buffer I and eluted as described earlier.

For the competition assay between Hrr25p and TRAPPI, 0.3 µM of GST–Sec23p or GST beads were incubated with 0.25 µM of purified TRAPPI and/or 0.15 µM of His₆–Hrr25p in binding buffer I for 3–4 h at 4 °C. For the competition assay between Sar1p and TRAPPI, 0.1 µM of GST–Sec23p or GST beads were incubated with 10 nM of purified His₆–Δ23Sar1p–GTPγS with increasing concentrations (0–50 nM) of TRAPPI as above. For the reciprocal experiment, GST fusion proteins were incubated with 10 nM of TRAPPI and increasing concentrations (0–50 nM) of His₆–Δ23Sar1p–GTPγS. The beads were then washed three times with binding buffer I and eluted as above. The molarity of TRAPPI was calculated based on the amount of Trs33p in the complex.

Yeast immunofluorescence microscopy. Cells expressing Vrg4p–GFP or Sec7p–GFP and Hrr25p–RFP were grown to an OD_{599 nm} of 0.5–1.5 in YPD medium. One to two OD_{599 nm} units were pelleted and resuspended in 25 µl of YPD medium. Cells were examined with a Carl Zeiss Observer Z.1 spinning-disk confocal fluorescence microscope using DIC, GFP, or RFP filters with a ×100 oil-immersion objective. Images were captured with a Zeiss AxioCam MRm and analysed using AxioVision Rel. 4.7 software. At least 300 puncta (and 100 cells) were examined in three separate experiments that were used to calculate the s.d. shown in Fig. 2a.

Growth conditions for the *hrr25* mutant. SFNY1941 (*MATa ura3-52 lys2-801 ade2-101 trpΔ63 his3-Δ200 leu2-Δ1 hrr25Δ1::loxP-kanMX-loxP pKK204(2µ pGAL-3HA-HRR25^{degron})*) was grown in YP–Raf–Gal (2% Raffinose, 0.5% Galactose) medium to OD_{599 nm} = 1–2. A total of 500 OD_{599 nm} units of cells were pelleted

under sterile conditions and shifted to either YP–Raf–Gal or YPD (YP + 2% glucose) medium for 10 h.

Differential centrifugation experiment. A total of 100 OD_{599 nm} units of cells were pelleted, resuspended in 2 ml of spheroplast buffer (1.4 M sorbitol, 100 mM sodium phosphate pH 7.5, 0.35% 2-mercaptoethanol and 0.5 mg ml^{−1} zymolyase) and incubated for 30 min at 37 °C. The spheroplasted cells were then divided into four 0.5 ml aliquots and centrifuged over a 1 ml sorbitol cushion (1.7 M sorbitol, 100 mM HEPES pH 7.2) for 5 min at 3,800g at 4 °C in a microfuge. The supernatant was removed and the four pellets were resuspended in 1 ml of lysis buffer (100 mM HEPES pH 7.2, 1 mM EGTA, 0.2 mM DTT, 1 mM PMSF and protease inhibitors) and lysed using a Dounce homogenizer. The lysate was centrifuged for 2 min at 500g at 4 °C in a microfuge and the supernatant was transferred to a new tube. An aliquot of this fraction (100 µl) was mixed with 50 µl of 3× sample buffer (total fraction, T) and heated at 100 °C for 5 min, while the remaining portion (600 µl) was centrifuged for 90 min at 190,000g at 4 °C in a Beckman SW55 Ti rotor. The lipid layer was removed and 100 µl of the supernatant (S) was mixed with 50 µl of 3× sample buffer and heated at 100 °C for 5 min. The pellet (P) was resuspended in 500 µl of lysis buffer and 100 µl was mixed with 50 µl of 3× sample buffer and heated at 100 °C for 5 min.

Immunoprecipitation assay to detect phosphorylation. SFNY1941 was grown as described earlier and whole-cell lysates (10 mg) were immunoprecipitated with anti-Sec24 antibody. The immunoprecipitates were then immunoblotted with anti-phospho-Ser/Thr (BD Biosciences, 1:500 dilution), anti-Sec24 (1:1,000 dilution) and anti-Sec23 antibodies (1:1,000 dilution).

Mass spectrometry analysis. *In vitro* phosphorylated GST–Sec23p was trypsin-digested and subjected to liquid chromatography coupled with tandem mass spectrometry (LC–MS/MS) analysis as described previously²³.

Generation of Sec23p phosphomimetic mutations. The phosphomimetic mutations (S742D, T747E and S742D/T747E) in GST–Sec23p were generated by the two-step PCR method for site-directed mutagenesis using pPE124 (ref. 24) as the template. Mutations in pRS414-SEC23(S742D/T747E) were generated on pCF364 (ref. 25) using the QuikChange Site-directed mutagenesis kit (Agilent Technologies). All constructs were verified by sequencing.

Analysis of mBet3 and mSec23 on mammalian COPII vesicles. A detailed description of the generation, immunoisolation and immunoblotting of COPII transport intermediates derived from semi-intact normal rat kidney (NRK) cells is described elsewhere¹⁷. Briefly, a VSV–G–Myc construct was introduced by electroporation and its expression in the ER was amplified using vaccinia virus VTF-7 at 41 °C. After permeabilization, the VSV–G–Myc-expressing cells or control untransfected NRK cells were suspended in a vesicle budding cocktail and incubated at 32 °C for 30 min. Subsequently, the donor cells were removed by sedimentation. For the p115 experiment, the supernatant (which contains released COPII vesicles) was then incubated with purified full-length His₆–p115 (ref. 26) for 60 min at 32 °C. The p115 preparation was functional in tests of interactions with ER–Golgi SNAREs (not shown). The suspension of transport intermediates was then subjected to immunoisolation using anti-Myc antibody. Proteins were eluted from the beads using 0.1 M glycine pH 2.5, neutralized, concentrated, and analysed on a 4–20% gradient SDS polyacrylamide gel followed by western blot analysis. To quantitate the abundance of mBet3 and mSec23 on the isolated vesicles and to normalize to vesicle yield, we divided the band intensity by the signal for the cargo marker VSV–G–Myc and syntaxin 5 for each lane. The cargo-normalized mBet3 and mSec23 signals were then expressed as a percentage relative to the GTP condition. GMP–PNP was used at 100 µM concentration.

Analysing ER–Golgi traffic *in vivo* in the presence of IC261. NRK cells were electroporated with a plasmid encoding VSV–G ts045–GFP, plated on glass coverslips in 6-well plates and incubated overnight at 40 °C. Ten minutes before the temperature shift, 100 µM IC261 (solubilized in DMSO) or DMSO was added to the medium at 40 °C. After 10 min, the coverslips were either fixed in 4% paraformaldehyde for 30 min, or shifted to 15 °C medium containing IC261 or DMSO. At 15 °C, VSV–G–GFP can leave the ER, but accumulates in swollen peripheral ER–Golgi interface structures that only slowly move towards the Golgi area²⁷. The 15 °C treatment makes ER exit, pre-Golgi assembly and transport to the Golgi more resolvable. Coverslips were fixed after 30 or 60 min at 15 °C.

After fixation, coverslips were treated twice for 10 min with 0.1 M glycine and then the samples were permeabilized in BSA/goat serum blocking solution containing 0.35% saponin. All subsequent antibody incubations and washes were carried out in blocking solution containing saponin. Primary antibody incubations included the anti-rbet1 mouse monoclonal antibody 16G6 (ref. 28) and a rabbit polyclonal antisera against gpp130 (Covance Research Products). 16G6 is known to label rbet1 more intensely when the antigen accumulates at peripheral sites²⁹. Secondary antibodies were goat anti-mouse-cy3 and goat anti-rabbit-cy5. DAPI was also included during the secondary antibody incubation. After extensive

washing, the coverslips were mounted in Slow-fade Gold mounting medium (Invitrogen) and imaged using the wide-field microscope and instrumentation methods described before²⁹. Briefly, each field of cells were captured in four colours (GFP, cy3, cy5 and DAPI) at 21 focal planes through the sample. Image stacks were then deconvolved using the Huygens algorithm (Scientific Volume Imaging). Maximum intensity projections from five consecutive image planes were used for quantification and display.

To quantify the transfer from the ER to pre-Golgi structures, each set of images was opened as an image stack in the Openlab program (Improvision). The extracellular background was subtracted from each image, and then a roughly square region of interest (ROI) that abutted the nucleus and extended approximately three-quarters of the distance to the edge of the cell was hand-drawn. This ROI was chosen such that it constituted a region of cytoplasm completely free of any Golgi labelling. The rbt1 image was used to define a pre-Golgi mask within this larger ROI using an intensity threshold of 4–6× the rbt1 labelling background. The pre-Golgi mask was then subtracted from the original ROI to create an ER mask. The pre-Golgi and ER mask were sequentially applied to the VSV-G–GFP image to determine the average maximum intensity of cargo spots in the pre-Golgi compartment, and the average intensity of cargo in the ER, respectively. The ratio of these two parameters, average maximum intensity in the pre-Golgi compartment divided by average intensity in the ER, was determined for each of the twenty randomly sampled images from each condition.

To quantify the concentration of pre-Golgi structures near the Golgi, a Golgi mask was derived from the gpp130 image using an intensity threshold of 10% of maximum intensity for the Golgi area of interest. A total punctate cargo mask was derived from the VSV-G–GFP image by choosing an intensity threshold for each cell sample such that a faint punctate spot would be captured in the mask but residual diffuse ER labelling would not. This cargo mask was then superimposed on the Golgi mask and all cargo-containing objects that did not partially or

completely overlap with a Golgi object were deleted from the cargo mask. This peri-Golgi cargo mask, as well as the original total cargo mask were sequentially applied to the VSV-G–GFP image to determine the total intensity of cargo spots in the peri-Golgi region and in the whole cell, respectively. The ratio of these two parameters, total intensity of cargo spots in the peri-Golgi region divided by total intensity of cargo spots in the cell, was determined for each of the 20 randomly sampled images analysed for each condition.

For both quantifications, the average raw ratio for DMSO-treated cells from a given experiment was defined as 100% and each individual ratio value from that experiment was expressed relative to 100%. This normalization step allowed combination of quantifications from independent experiments to produce the values for Fig. 5c and e.

All experiments in the manuscript were performed at least three times or more on separate days.

23. Guttman, M. *et al.* Interactions of the NPXY microdomains of the low density lipoprotein receptor-related protein 1. *Proteomics* **9**, 5016–5028 (2009).
24. Gimeno, R. E., Espenshade, P. & Kaiser, C. A. COPII coat subunit interactions: Sec24p and Sec23p bind to adjacent regions of Sec16p. *Mol. Biol. Cell* **7**, 1815–1823 (1996).
25. Fromme, J. C. *et al.* The genetic basis of a craniofacial disease provides insight into COPII coat assembly. *Dev. Cell* **13**, 623–634 (2007).
26. Diao, A., Frost, L., Morohashi, Y. & Lowe, M. Coordination of golgin tethering and SNARE assembly: GM130 binds syntaxin 5 in a p115-regulated manner. *J. Biol. Chem.* **283**, 6957–6967 (2008).
27. Saraste, J. & Svensson, K. Distribution of the intermediate elements operating in ER to Golgi transport. *J. Cell Sci.* **100**, 415–430 (1991).
28. Hay, J. C. *et al.* Localization, dynamics, and protein interactions reveal distinct roles for ER and Golgi SNAREs. *J. Cell Biol.* **141**, 1489–1502 (1998).
29. Bentley, M. *et al.* Vesicular calcium regulates coat retention, fusogenicity, and size of pre-Golgi intermediates. *Mol. Biol. Cell* **21**, 1033–1046 (2010).

Hot Jupiters from secular planet–planet interactions

Smadar Naoz¹, Will M. Farr¹, Yoram Lithwick¹, Frederic A. Rasio¹ & Jean Teyssandier¹

About 25 per cent of ‘hot Jupiters’ (extrasolar Jovian-mass planets with close-in orbits) are actually orbiting counter to the spin direction of the star¹. Perturbations from a distant binary star companion^{2,3} can produce high inclinations, but cannot explain orbits that are retrograde with respect to the total angular momentum of the system. Such orbits in a stellar context can be produced through secular (that is, long term) perturbations in hierarchical triple-star systems. Here we report a similar analysis of planetary bodies, including both octupole-order effects and tidal friction, and find that we can produce hot Jupiters in orbits that are retrograde with respect to the total angular momentum. With distant stellar mass perturbers, such an outcome is not possible^{2,3}. With planetary perturbers, the inner orbit’s angular momentum component parallel to the total angular momentum need not be constant⁴. In fact, as we show here, it can even change sign, leading to a retrograde orbit. A brief excursion to very high eccentricity during the chaotic evolution of the inner orbit allows planet–star tidal interactions to rapidly circularize that orbit, decoupling the planets and forming a retrograde hot Jupiter.

Despite many attempts^{2,3,5–11}, there is no model that can account for all the properties of the known hot Jupiter systems. One model suggests that hot Jupiters formed far away from the star and slowly spiralled in, losing angular momentum and orbital energy to the protoplanetary disk^{12,13}. This ‘migration’ process should produce planets with low orbital inclinations and eccentricities. However, many hot Jupiters are observed to be on orbits with high eccentricities and misaligned with the spin axis of the star (as measured through the Rossiter–McLaughlin effect¹⁴), and some of these (8 out of 32) even appear to be orbiting counter to the spin of the star. In a second model, secular perturbations from a distant binary star companion can produce increases in the eccentricity and inclination of a planetary orbit¹⁵. During the evolution to high eccentricity, tidal dissipation near pericentre can force the planet’s orbit to decay, potentially forming a misaligned hot Jupiter^{2,3}. Recently, secular chaos involving several planets has also been proposed as a way to form hot Jupiters on eccentric and misaligned orbits¹¹. A different class of models that can produce a tilted orbit involves planet–planet scattering⁵, possibly combined with other perturbers and tidal friction⁷. In such models, the initial configuration is a densely packed system of planets and the final tilted orbit is a result of dynamical scattering among the planets, in contrast to the secular interactions we study here.

In our general treatment of secular interactions between two orbiting bodies, we allow the magnitude and orientation of both orbital angular momenta to change (Fig. 1). The outer body (here either a planet or a brown dwarf) gravitationally perturbs the inner planet on timescales long compared to both orbital periods (that is, we consider the secular evolution of the system). We define the orientation of the inner orbit with respect to the invariable plane of the system (the invariable plane is perpendicular to the total angular momentum): a prograde (retrograde) orbit has $i_1 < 90^\circ$ ($i_1 > 90^\circ$), where i_1 is the angle between the inner orbit’s orbital angular momentum vector and the total angular momentum vector. Note that the word ‘retrograde’ is also used in the literature to indicate orbital motion counter to the stellar spin. The directly observed parameter is actually the projected angle between

the spin axis of the star and the orbital angular momentum of a hot Jupiter. Our proposed mechanism can produce hot Jupiters that are ‘retrograde’ with respect to both the stellar spin and the total angular momentum. By contrast, a distant stellar companion can only succeed in the former (see Supplementary Information for details): here we will use the term ‘retrograde’ in only one sense, that is, to indicate an orbit with $i_1 > 90^\circ$, as defined above.

We assume a hierarchical configuration, with the outer perturber on a much wider orbit than the inner one. In the secular approximation,

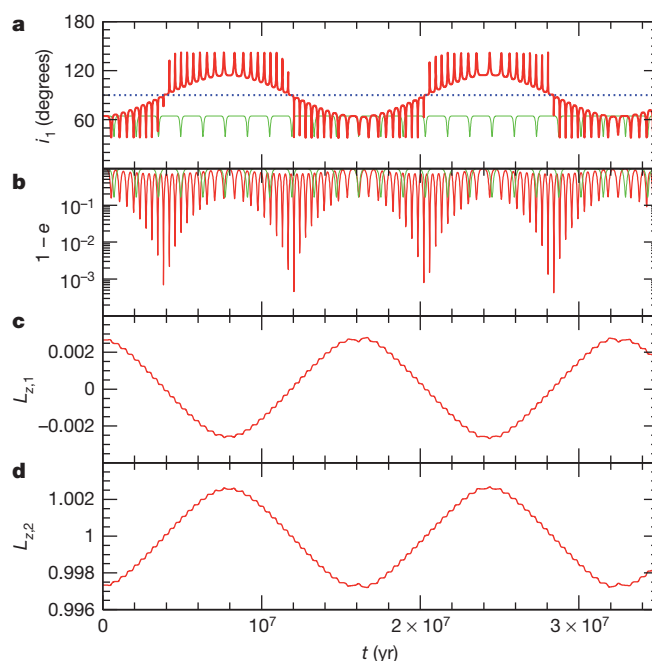


Figure 1 | Dynamical evolution of a representative planet and brown dwarf system. Here we ignore tidal dissipation, but we do include the lowest-order post-Newtonian precession rate for the inner orbit. The star has mass $1M_\odot$, the planet has mass $1M_J$ and the outer brown dwarf has mass $40M_J$. The inner orbit has $a_1 = 6$ AU and the outer orbit has $a_2 = 100$ AU. The initial eccentricities are $e_1 = 0.001$ and $e_2 = 0.6$, and the initial relative inclination is $i = 65^\circ$. Red curves show the following: **a**, the inner orbit’s inclination, i_1 ; **b**, the eccentricity of the inner orbit (as $1 - e_1$); and **c**, **d**, the z -component of the angular momentum of the inner ($L_{z,1}$; **c**) and outer ($L_{z,2}$; **d**) orbit, normalized to the total angular momentum (where the z -axis is defined to be along the total angular momentum). The dotted line in **a** marks the 90° boundary, separating prograde and retrograde orbits. The initial mutual inclination of 65° corresponds to an inner and outer inclination with respect to the total angular momentum (parallel to z) of 64.7° and 0.3° , respectively. During the evolution, the eccentricity and inclination of the inner orbit oscillate, but, in contrast to what would be predicted from evolution equations truncated to quadrupole order (shown by green curves in **a** and **b**), the eccentricity of the inner orbit can occasionally reach extremely high values and its inclination can become higher than 90° . The outer orbit’s inclination always remains near its initial value. We note that more compact systems usually do not exhibit the same kind of regular oscillations between retrograde and prograde orbits illustrated here, as chaotic effects become more important and are revealed at octupole order (see Fig. 2). We find that for $\sim 50\%$ of the time, the inner orbit is retrograde.

¹Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA), Northwestern University, Evanston, Illinois 60208, USA.

the orbits may change shape and orientation but the semi-major axes are strictly conserved in the absence of tidal dissipation^{4,16}. In particular, the Kozai–Lidov mechanism^{17–19} produces large-amplitude oscillations of the eccentricity and inclination when the initial relative inclination between the inner and outer orbits is sufficiently large ($40^\circ < i < 140^\circ$).

We have derived the secular evolution equations to octupole order using Hamiltonian perturbation theory^{4,20,21}. In contrast to previous derivations of ‘Kozai-type’ evolution, our treatment allows for changes in the z -components of both orbital angular momenta (that is, the components along the total angular momentum), $L_{z,1}$ and $L_{z,2}$, where 1 and 2 refer to the inner and outer orbits, respectively (see Supplementary Information). The octupole-order equations allow us to calculate the evolution of systems with more closely coupled orbits and with planetary-mass perturbers. The octupole-level terms can give rise to fluctuations in the eccentricity maxima to arbitrarily high values^{4,21}, in

contrast to the regular evolution in the quadrupole potential^{2,3,19}, where the amplitude of eccentricity oscillations is constant.

Many previous studies of secular perturbations in hierarchical triples considered a stellar-mass perturber, for which $L_{z,1}$ is very nearly constant^{2,3,19}. Moreover, the assumption that $L_{z,1}$ is constant has been built into previous derivations^{22,23,24}. However, this assumption is only valid as long as $L_2 \gg L_1$, which is not the case in comparable-mass systems (for example, with two planets). Unfortunately, an immediate consequence of this assumption is that an orbit that is prograde relative to the total angular momentum always remains prograde. Figure 1 shows the evolution of a representative system (here without tidal effects for simplicity): the inner planet oscillates between prograde and retrograde orbits (with respect to the total angular momentum) as angular momentum flows back and forth between the two orbits.

Previous calculations of planet migration through ‘Kozai cycles with tidal friction’^{2,3,16,19} produced a slow, gradual spiral-in of the inner planet.

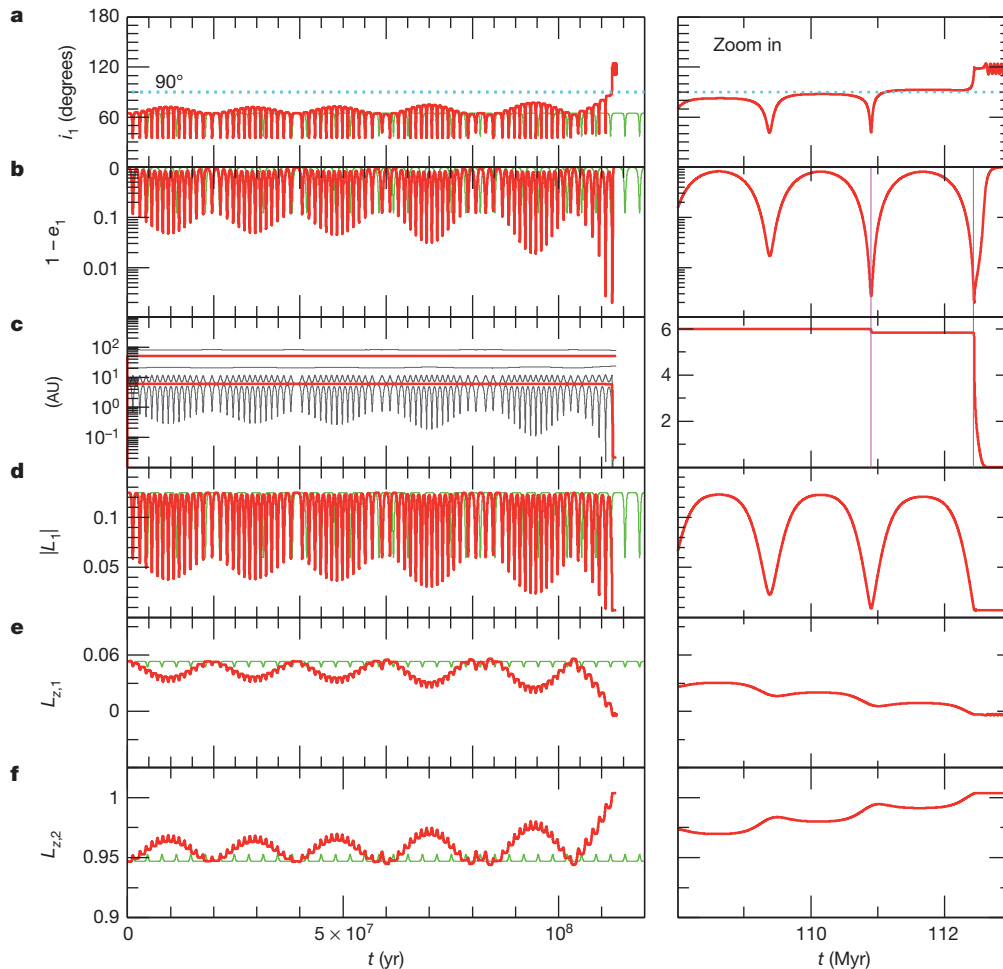


Figure 2 | Dynamical evolution of a representative two-planet system with tidal dissipation included. The inner planet becomes retrograde at 112 Myr, and remains retrograde after circularizing into a hot Jupiter. Here the star has mass $1M_\odot$, the inner planet has mass $1M_J$ and the outer planet has mass $3M_J$. The inner orbit has $a_1 = 6$ AU and the outer orbit has $a_2 = 61$ AU. The initial eccentricities are $e_1 = 0.01$ and $e_2 = 0.6$, the initial relative inclination $i = 71.5^\circ$, and the argument of periapsis is 45° . Left panels, complete simulation; right panels, zoomed-in view around time $t \approx 110$ Myr. Red curves show: **a**, the inner orbit’s inclination (i_1); **b**, the eccentricity of the inner orbit (as $1 - e_1$); **c**, the semi-major axis for the inner orbit and the outer orbit; **d**, the magnitude of the angular momentum of the inner orbit; and in **e** and **f**, the z -component of the angular momentum of the inner ($L_{z,1}$; **e**) and outer ($L_{z,2}$; **f**) orbit, normalized to the total angular momentum. The black curves in **c** are the pericentre and apocentre distances of the inner and outer orbits, respectively. The initial mutual inclination of 71.5° corresponds to inner- and outer-orbit inclinations

of 64.7° and 6.8° , respectively. During each excursion to very high eccentricity for the inner orbit (marked with vertical lines in **b** and **c** right panels), tidal dissipation becomes significant. Eventually the inner planet is tidally captured by the star and its orbit becomes decoupled from the outer body. After this point, the orbital angular momenta remain nearly constant. The final semi-major axis for the inner planet is 0.022 AU, typical of a hot Jupiter. The green curves in **a**, **b**, **d**, **e** and **f** show the evolution in the quadrupole approximation (but including tidal friction), demonstrating that the octupole-order effects lead to a qualitatively different behaviour. For the tidal evolution in this example, we assume tidal quality factors $Q_* = 5.5 \times 10^6$ for the star and $Q_J = 5.8 \times 10^6$ for the hot Jupiter (see Supplementary Information). We monitor the pericentre distance of the inner planet to ensure that it always remains outside the Roche limit²⁹. Here, as in Fig. 1, we also include the lowest-order post-Newtonian precession rate for the inner orbit.

Instead, our treatment shows that the eccentricity can occasionally reach a much higher value than in the regular ‘Kozai cycles’ calculated to quadrupole order. Thus, the pericentre distance will occasionally shrink on a short timescale (compared to the Kozai period), and the planet can then suddenly be ‘tidally captured’ (that is, the inner orbit suddenly decouples from the outer one, and undergoes rapid tidal circularization) by the star. We propose to call this ‘Kozai capture’.

Kozai capture provides a new way to form hot Jupiters. If the capture happens after the inner orbit has flipped, the hot Jupiter will appear in a retrograde orbit. This is illustrated in Fig. 2. During the evolution of the system, the inner orbit shrinks in steps (Fig. 2c) whenever the dissipation becomes significant, that is, near unusually high eccentricity maxima. The inner orbit can then eventually become tidally circularized. This happens near the end of the evolution, on a very short timescale (see Fig. 2, right panels). In this final step, the inner orbit completely and quickly decouples from the outer perturber, and the orbital angular momenta then become constant. Therefore, the final semi-major axis for the hot Jupiter is $\sim 2r_p$, where r_p is the pericentre distance at the beginning of the capture phase²⁵.

The same type of evolution shown in Fig. 2 is seen with a broad range of initial conditions. There are two main routes to forming a hot Jupiter through the dynamical evolution of the systems we consider here. In the first, tidal friction slowly damps the growing eccentricity of the inner planet, resulting in a circularized, prograde hot Jupiter. In the second, a sudden high-eccentricity spike in the orbital evolution of the inner planet is accompanied by a flip of its orbit. The planet is then quickly circularized into a retrograde short-period orbit. We can estimate the relative frequencies of these two types of outcomes using Monte Carlo simulations. Given the vast parameter space for initial conditions, a complete study of the statistics is beyond the scope of this Letter (but see S.N. *et al.*, manuscript in preparation). However, we can provide a representative example: consider systems where the inner planet was formed *in situ* at a distance from the star of $a_1 = 6$ AU with zero obliquity (that is, it orbits in the stellar equatorial plane) and with some small eccentricity $e_1 = 0.01$, while the outer planet has $a_2 = 61$ AU. The masses are $m_1 = 1M_J$ and $m_2 = 3M_J$ (M_J is the Jovian mass). We draw the eccentricity of the outer orbit from a uniform distribution and the mutual inclination from a distribution uniform in $\cos i$ between 0 and 1 (that is, isotropic orbits among prograde ones). For this case we find that, among all hot Jupiters that are formed, about 7% are in truly retrograde motion (that is, with respect to the total angular momentum) and about 50% are orbiting counter to the stellar spin direction. Note that the latter fraction is significantly larger than the values previous studies have obtained with stellar-mass perturbers (at most $\sim 10\%$; refs 2, 3). The high observed incidence of planets orbiting counter to the stellar spin axis¹ may suggest that planet–planet secular interactions are an important part of their dynamical history.

Our mechanism requires that two coupled orbits start with a relatively high mutual inclination ($i > 50^\circ$). In Fig. 2, the outer planet is initially on a very wide orbit, similar to those of directly imaged planets such as Fomalhaut b (ref. 26) and HR 8799b (ref. 27). The inner planet’s initial location is consistent with *in situ* formation on a nearly circular orbit, in accordance with the standard core accretion model²⁸. An alternative path to such a configuration involves strong planet–planet scattering in a closely packed initial system of several giant planets⁷. Independent of any particular planet formation mechanism, we predict that systems with misaligned hot Jupiters should also contain a much more distant massive planet or brown dwarf on an inclined orbit.

Received 23 October 2010; accepted 24 March 2011.

1. Triaud, A. H. M. J. *et al.* Spin-orbit angle measurements for six southern transiting planets. New insights into the dynamical origins of hot Jupiters. *Astron. Astrophys.* **524**, A25 (2010).

2. Fabrycky, D. & Tremaine, S. Shrinking binary and planetary orbits by Kozai cycles with tidal friction. *Astrophys. J.* **669**, 1298–1315 (2007).
3. Wu, Y., Murray, N. W. & Ramsahai, J. M. Hot Jupiters in binary star systems. *Astrophys. J.* **670**, 820–825 (2007).
4. Ford, E. B., Kozinsky, B. & Rasio, F. A. Secular evolution of hierarchical triple star systems. *Astrophys. J.* **535**, 385–401 (2000).
5. Chatterjee, S., Matsumura, S., Ford, E. B. & Rasio, F. A. Dynamical outcomes of planet–planet scattering. *Astrophys. J.* **686**, 580–602 (2008).
6. Lai, D., Foucart, F. & Lin, D. N. C. Evolution of spin direction of accreting magnetic protostars and spin-orbit misalignment in exoplanetary systems. *Mon. Not. R. Astron. Soc.* (submitted); preprint at (<http://arxiv.org/abs/1008.3148>) (2011).
7. Nagasawa, M., Ida, S. & Bessho, T. Formation of hot planets by a combination of planet scattering, tidal circularization, and the Kozai mechanism. *Astrophys. J.* **678**, 498–508 (2008).
8. Schlaufman, K. C. Evidence of possible spin-orbit misalignment along the line of sight in transiting exoplanet systems. *Astrophys. J.* **719**, 602–611 (2010).
9. Takeda, G., Kita, R. & Rasio, F. A. Planetary systems in binaries. I. Dynamical classification. *Astrophys. J.* **683**, 1063–1075 (2008).
10. Winn, J. N., Fabrycky, D., Albrecht, S. & Johnson, J. A. Hot stars with hot Jupiters have high obliquities. *Astrophys. J.* **718**, L145–L149 (2010).
11. Wu, Y. & Lithwick, Y. Secular chaos and the production of hot Jupiters. Preprint at (<http://arxiv.org/abs/1012.3475>) (2010).
12. Lin, D. N. C. & Papaloizou, J. On the tidal interaction between protoplanets and the proto-planetary disk. III — Orbital migration of protoplanets. *Astrophys. J.* **309**, 846–857 (1986).
13. Masset, F. S. & Papaloizou, J. Runaway migration and the formation of hot Jupiters. *Astrophys. J.* **588**, 494–508 (2003).
14. Gaudi, B. S. & Winn, J. N. Prospects for the characterization and confirmation of transiting exoplanets via the Rossiter-McLaughlin effect. *Astrophys. J.* **655**, 550–563 (2007).
15. Holman, M., Touma, J. & Tremaine, S. Chaotic variations in the eccentricity of the planet orbiting 16 Cygni B. *Nature* **386**, 254–256 (1997).
16. Eggleton, P. P., Kiseleva, L. G. & Hut, P. The equilibrium tide model for tidal friction. *Astrophys. J.* **499**, 853–870 (1998).
17. Kozai, Y. Secular perturbations of asteroids with high inclination and eccentricity. *Astron. J.* **67**, 591–598 (1962).
18. Lidov, M. L. The evolution of orbits of artificial satellites of planets under the action of gravitational perturbations of external bodies. *Planet. Space Sci.* **9**, 719–759 (1962).
19. Mazeh, T. & Shaham, J. The orbital evolution of close triple systems — the binary eccentricity. *Astron. Astrophys.* **77**, 145–151 (1979).
20. Harrington, R. S. The stellar three-body problem. *Celest. Mech.* **1**, 200–209 (1969).
21. Krymowski, Y. & Mazeh, T. Studies of multiple stellar systems — II. Second-order averaged Hamiltonian to follow long-term orbital modulations of hierarchical triple systems. *Mon. Not. R. Astron. Soc.* **304**, 720–732 (1999).
22. Kiseleva, L. G., Eggleton, P. P. & Mikkola, S. Tidal friction in triple stars. *Mon. Not. R. Astron. Soc.* **300**, 292–302 (1998).
23. Zdziarski, A. A., Wen, L. & Gierliński, M. The superorbital variability and triple nature of the X-ray source 4U 1820–303. *Mon. Not. R. Astron. Soc.* **377**, 1006–1016 (2007).
24. Mikkola, S. & Tanikawa, K. Does Kozai resonance drive CH Cygni? *Astron. J.* **116**, 444–450 (1998).
25. Ford, E. B. & Rasio, F. A. On the relation between hot Jupiters and the Roche limit. *Astron. J.* **638**, L45–L48 (2006).
26. Kalas, P. *et al.* Optical images of an exosolar planet 25 light-years from Earth. *Science* **322**, 1345–1348 (2008).
27. Marois, C. *et al.* Direct imaging of multiple planets orbiting the star HR 8799. *Science* **322**, 1348–1352 (2008).
28. Pollack, J. B. *et al.* Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **124**, 62–85 (1996).
29. Matsumura, S., Peale, S. J. & Rasio, F. A. Formation and evolution of close-in planets. *Astrophys. J.* **725**, 1995–2016 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Fabrycky and H. Perets for discussions. S.N. acknowledges support from a Gruber Foundation Fellowship and from the National Post Doctoral Award Program for Advancing Women in Science (Weizmann Institute of Science). Simulations for this project were performed on the HPC cluster *fugu* funded by an NSF MRI award.

Author Contributions S.N. performed numerical calculations with help from J.T. All authors developed the mathematical model, discussed the physical interpretation of the results and jointly wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.N. (snaz@northwestern.edu).

A single-atom quantum memory

Holger P. Specht^{1†}, Christian Nölleke¹, Andreas Reiserer¹, Manuel Uphoff¹, Eden Figueroa¹, Stephan Ritter¹ & Gerhard Rempe¹

The faithful storage of a quantum bit (qubit) of light is essential for long-distance quantum communication, quantum networking and distributed quantum computing¹. The required optical quantum memory must be able to receive and recreate the photonic qubit; additionally, it must store an unknown quantum state of light better than any classical device. So far, these two requirements have been met only by ensembles of material particles that store the information in collective excitations^{2–7}. Recent developments, however, have paved the way for an approach in which the information exchange occurs between single quanta of light and matter^{8–13}. This single-particle approach allows the material qubit to be addressed, which has fundamental advantages for realistic implementations. First, it enables a heralding mechanism that signals the successful storage of a photon by means of state detection^{14–16}; this can be used to combat inevitable losses and finite efficiencies. Second, it allows for individual qubit manipulations, opening up avenues for *in situ* processing of the stored quantum information. Here we demonstrate the most fundamental implementation of such a quantum memory, by mapping arbitrary polarization states of light into and out of a single atom trapped inside an optical cavity. The memory performance is tested with weak coherent pulses and analysed using full quantum process tomography. The average fidelity is measured to be 93%, and low decoherence rates result in qubit coherence times exceeding 180 microseconds. This makes our system a versatile quantum node with excellent prospects for applications in optical quantum gates¹⁷ and quantum repeaters¹⁸.

The efficient and reversible exchange of quantum information between single flying qubits (photons) and single quantum emitters (material particles) is a challenge, as it requires a sufficiently large interaction strength. A promising approach to this goal is to trap the emitter inside an optical cavity, thereby providing strong coupling between light and matter. Within this setting, optical control has been achieved, although unidirectionally, by mapping quantum information from a single atom onto a photon¹¹. The reverse process of transferring quantum information from a light field onto an atom has proved to be much more elusive and far more difficult to achieve. Experimentally, the problem has been addressed only recently, by sending a weak laser pulse onto the cavity and controlling its transmission through the cavity in two ways: by means of electromagnetically induced transparency^{19,20} and by storing part of the pulse in the atom and releasing it¹².

In order to advance this approach to the realization of a quantum memory for photonic qubits, robustness against inevitable photon loss during propagation and detection becomes essential. Our implementation uses qubits encoded in the polarization degree of freedom such that photon loss reduces the efficiency of the quantum protocol, but does not compromise its fidelity. When compared to all previous demonstrations of optical quantum memories^{2–7}, our system has competitive fidelity and efficiency, and longer qubit coherence times. Even better, clear strategies for improving its characteristics are identified. No fundamental physical limits have been reached, testifying to the usefulness of this approach and indicating a promising future for single-atom quantum memories.

The hardware of our quantum memory is a single ⁸⁷Rb atom quasi-permanently trapped at the centre of a high-finesse optical cavity (see Fig. 1 and Supplementary Information)¹⁹. The efficiencies for coupling photons into and out of the cavity are optimized by using one high-reflectivity mirror (transmission $T < 6$ p.p.m.) and an output coupler with higher transmission ($T \approx 100$ p.p.m.). The atom-cavity coupling constant g is comparable to the cavity-field decay rate κ and the atomic polarization decay rate of ⁸⁷Rb (γ), and therefore the system is in the intermediate coupling regime of cavity quantum electrodynamics: $(g, \kappa, \gamma) = 2\pi \times (5, 2.5, 3)$ MHz.

The memory is initialized by preparing the atom in the state $|F = 1, m_F = 0\rangle$ with an efficiency higher than 90%. It is probed using weak coherent pulses (mean photon number $\bar{n} < 1$) that are resonant with the cavity and are detuned by 12 MHz to the blue from the Stark-shifted $F = 1 \leftrightarrow F' = 1$ resonance of the D₂ line. Simultaneously with the arrival of the incoming pulse, the Rabi frequency of a control laser in Raman resonance with the cavity is adiabatically ramped down to zero. The control laser is oriented perpendicular and π -polarized with respect to the quantization axis defined by the cavity axis. This transfers the atomic population to the ground states $|F = 2, m_F = \pm 1\rangle$ through a cavity-mediated stimulated Raman adiabatic passage^{8,21} (Fig. 1a). During this coherent process, the phase relation between the σ^{\pm} input polarization modes is mapped to a relative phase between the populations of the above-mentioned Zeeman substates (Fig. 1b). Efficient storage is achieved by optimizing the centre and width of the falling edge of the control laser pulse, which follows a \cos^2 function in time.

After a variable storage time, a second control laser pulse converts the atomic qubit back onto the polarization of a single photon¹¹; this ideally will have the same polarization state as the initial incoming pulse (Fig. 1c). The temporal shape of the retrieved photon's wave packet can thereby be adjusted^{22,23}, independent of the form of the input pulse, which is of great importance in more complex quantum networks. A non-polarizing beam splitter is used to separate the paths of incoming and outgoing photons, with the latter being directed to a detection set-up, where the polarization state of the photon can be analysed in any basis.

We define the efficiency of our quantum memory as the fraction of photons retrieved from the cavity after storage, normalized to the input (Fig. 2). It was measured to be $(9.3 \pm 1)\%$ (Supplementary Information). The uncertainty is dominated by fluctuations in the system parameters (for example, the atomic position). We independently measured the photon production efficiency of the readout process to be 56%. The efficiency is clearly limited by the atom-cavity coupling^{19,24} and could be increased considerably by better localization of the atom or a smaller cavity mode volume. Owing to the single-atom character of our quantum memory, only single photons can be stored, even if the input is a coherent state. This is confirmed by the measured antibunching of the retrieved photons (Fig. 2 inset).

We have characterized our quantum memory with six different input polarizations: $|R\rangle$, $|L\rangle$, $|H\rangle$, $|V\rangle = (|R\rangle \pm |L\rangle)/\sqrt{2}$ and $|D\rangle$, $|A\rangle = (|R\rangle \pm i|L\rangle)/\sqrt{2}$. For each input state, the output is analysed in three orthogonal bases, which allows for a full reconstruction of the retrieved photon's density matrix²⁵. The fidelity is defined as the

¹Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Strasse 1, 85748 Garching, Germany. [†]Present address: Osram Opto Semiconductors GmbH, Leibnizstrasse 4, 93055 Regensburg, Germany.

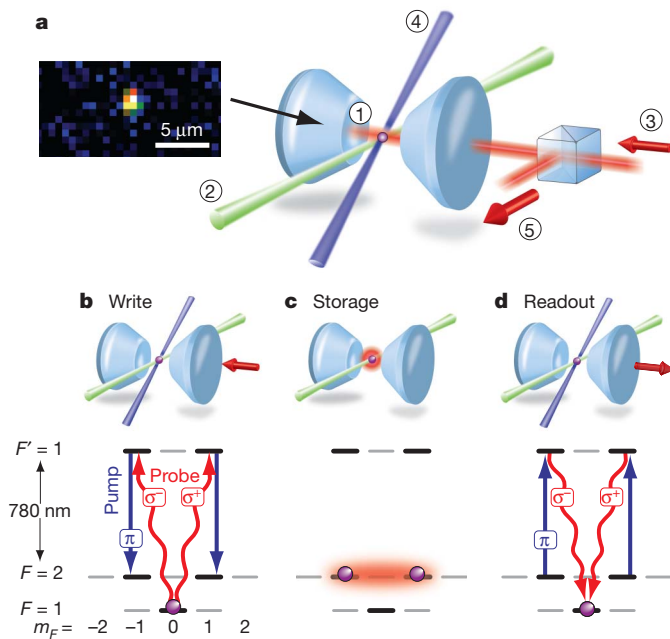


Figure 1 | Single-atom quantum memory. **a**, A single atom (1; purple sphere) is trapped at the centre of a high-finesse optical cavity (formed by the coned mirrors in blue) using a far-detuned standing-wave dipole trap (2). Inset, a typical fluorescence image of a trapped atom. **b–d**, Experimental set-up and atomic level diagram. **b**, An impinging weak coherent probe pulse (3 in **a**, and red curly arrows in **b**) with arbitrary polarization (given by a superposition of σ^+ and σ^- polarization) is converted into an atomic spin excitation using a π -polarized pump laser (4 in **a**, and blue straight arrows in **b** and **d**). **c**, The storage process maps the photonic polarization qubit onto a long-lived superposition of the $|F = 2, m_F = \pm 1\rangle$ ground states of the atom. The purple spheres and the red cloud signify the atomic population and a coherent superposition, respectively. **d**, After a variable storage time, the polarization qubit is retrieved by the production of a single photon (5 in **a** and red curly arrows in **d**).

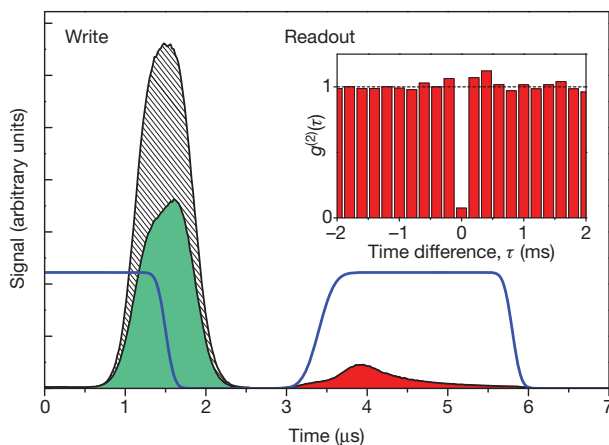


Figure 2 | Write and read processes of the memory. While the input photon pulse (hatched area, full-width at half-maximum 0.7 μ s) impinges on the cavity, the power of the control laser (blue, curve only schematic) is adiabatically ramped down to zero. Part of the incident light is directly reflected from the memory (green area). After an adjustable storage time, a photon is produced by adiabatically switching on the control laser. The area of the retrieved photon pulse (red) relative to the incident pulse yields the overall storage efficiency of the memory of 9.3%. Inset, correlation function $g^{(2)}(\tau)$ of the retrieved photons as measured by a Hanbury Brown–Twiss experiment. The ratio of two-photon to single-photon events is 0.5% and can be fully explained by stray light and dark counts of the detectors. This verifies the single-photon character of the retrieved light.

overlap of the density matrix ρ of the measured output with the ideal input state $|\psi_i\rangle$: $F = \langle \psi_i | \rho | \psi_i \rangle$. It therefore not only includes the storage process but also possible imperfections of the read-out. The fidelities of the six input states for a storage time of 2 μ s are as follows:

$$|H\rangle: 92.2(4)\%; |V\rangle: 92.0(4)\%;$$

$$|D\rangle: 91.9(5)\%; |A\rangle: 90.9(4)\%;$$

$$|R\rangle: 95.1(4)\%; |L\rangle: 94.2(4)\%;$$

yielding an arithmetic mean²⁶ of 92.7(2)%. The average fidelity of the input states directly reflected from the cavity is 99.20(2)%, proving that errors during preparation, propagation and detection of the photon's polarization state are small.

The fidelity we obtain has to be compared to what can be achieved following classical strategies. Unlike a quantum memory, a classical memory effectively performs a measurement on the state of the incoming qubit. On the basis of the incomplete information gained, it will try to reproduce the input state, a procedure known as an 'intercept-resend attack' in the context of quantum key distribution²⁷. The maximum average fidelity that can be achieved with such a classical memory and single input photons is 2/3 (ref. 28). When tested with coherent light, a classical memory could take advantage of pulses with higher photon number and gain additional information by measuring individual photons in different bases. It could also exploit non-unity efficiencies as a means to increase its fidelity by only emitting a photon when the incident photon number is high enough. Nevertheless, a quantum memory can also be characterized unambiguously with coherent input pulses, as long as the mean photon number in the input pulse and the efficiency of the memory are properly accounted for. In our case of 9.3% efficiency and a mean photon number of $\bar{n} = 1$, the fidelity threshold for a quantum memory is increased to 80% (see ref. 27 and Supplementary Information). The fidelity we obtain is well above this threshold, thereby proving the quantum nature of the memory.

In order to calculate the storage fidelity for an arbitrary input state, the process can be further analysed using quantum process tomography²⁹. Every polarization state can be represented by a three-dimensional vector \mathbf{S} . Pure states have $|\mathbf{S}| = 1$ and therefore lie on the unit Poincaré sphere. An intuitive representation for the imperfection of the storage process for any input state is given by the deformation of this sphere (Fig. 3a).

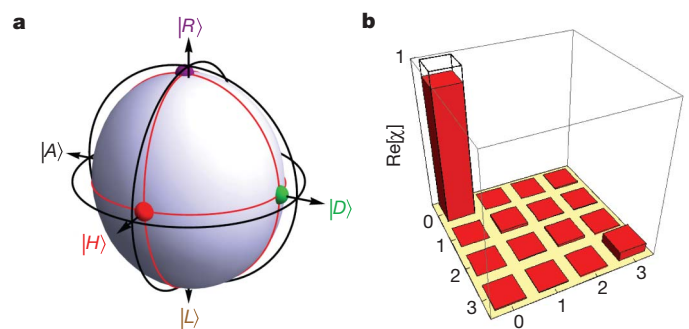


Figure 3 | Tomography of the storage process for a storage time of 2 μ s. **a**, The storage process can be depicted as a deformation (bluish ellipsoid) of the unit Poincaré sphere (black circles). The coloured dots indicate the polarizations of the input states on the axes (same colour coding) after storage. The average fidelity is $F = 92.7 \pm 0.2\%$, clearly proving the quantum character of the memory. **b**, Real part of the process matrix χ as obtained from a full quantum process tomography. All imaginary parts are close to zero (largest magnitude 0.034).

An alternative representation for the quantum process is obtained via the process matrix χ . It maps an input density matrix ρ_{in} onto the corresponding output state ρ_{out} via:

$$\rho_{\text{out}} = \sum_{m,n=0}^3 \chi_{mn} \sigma_m \rho_{\text{in}} \sigma_n^\dagger$$

Here, the operators σ_i form a set of Pauli matrices (\dagger denotes the adjoint operator). As we observe no systematic dependence of the efficiency on the polarization of the input pulse, the efficiency can be excluded by normalizing the density matrices. Then, for a process preserving the input state, χ_{00} is equal to one, while all other elements are zero. As shown in Fig. 3b, the main deviation of our χ matrix from the ideal one is a non-zero χ_{33} , indicating dephasing between the Zeeman states.

Several processes reduce the fidelity of our quantum memory. The main contribution arises from non-optimal optical pumping to $F = 1$, which leads to the production of a photon independent of the state of the input photon. By measuring the efficiency of producing a photon without an incoming photon during the storage process, we have confirmed that this undesired photon production occurs in 1.3% of the trials, which leads to an estimated reduction of the fidelity by 7%. The detrimental effect of incoherent transfer of the atomic population¹² is minimized by working in two-photon resonance but 12 MHz off-resonance from the atomic transition. Stray light and dark counts only occur with a probability of 0.3% relative to the number of incoming photons.

We have also characterized our memory by measuring the write-read fidelity as a function of the storage time (Fig. 4a). The classical limit of $2/3$ is reached after 82 μs . For longer storage times, the system approaches a fidelity of $F = 50\%$, as expected for a completely mixed output state. Two main processes contribute to the observed decay: residual fluctuating magnetic fields and differential Stark shifts caused by the trapping lasers. Both lead to a dephasing of the stored state. Gaussian field fluctuations explain the observed Gaussian decay of fidelity with storage time (Supplementary Information). We find that the fidelity for circularly polarized light decays slower than for linear polarizations, as the respective atomic states are eigenstates of any magnetic field along the quantization axis.

By applying a guiding magnetic field B_{guide} along this axis, the storage times can be further increased (Fig. 4b), because the guiding field suppresses the influence of perpendicular fluctuations ΔB on the absolute field stability by a factor $\Delta B/2B_{\text{guide}}$. The influence of longitudinal fluctuations, however, is unaffected. The guiding field causes the relative phase between the states $|F = 2, m_F = \pm 1\rangle$ to evolve at twice the Larmor frequency. As linear polarizations are stored as superpositions of these states, the measured fidelity shows a sinusoidal oscillation. Knowing the precession period of 21 μs for a guiding field of $B_{\text{guide}} = 34$ mG and the storage time, this can be compensated using polarization optics. The slightly reduced fidelity at short storage times is caused by Larmor precession that is already significant during the storage and retrieval processes. The overall result of applying a guiding field, however, is increased fidelity for a given storage time and therefore longer overall coherence times. This is evidenced by a more than twofold increase in coherence time, with the average fidelity reaching the classical limit of $2/3$ after 184 μs . The effect for purely circular polarizations, which are stored in eigenstates of the guiding field, is even more dramatic, with the fidelity staying almost constant during the observation period.

In summary, we have demonstrated a quantum memory for polarization qubits, based on a single trapped atom; this memory has an average fidelity of 93%, qubit coherence times of 184 μs and an overall efficiency of 9.3%. Longer storage times are possible by encoding the qubit in magnetic-field-insensitive states (clock states)^{30–32}. The single-particle nature of our memory also provides the potential for the implementation of a herald. As a successful storage attempt depopulates the

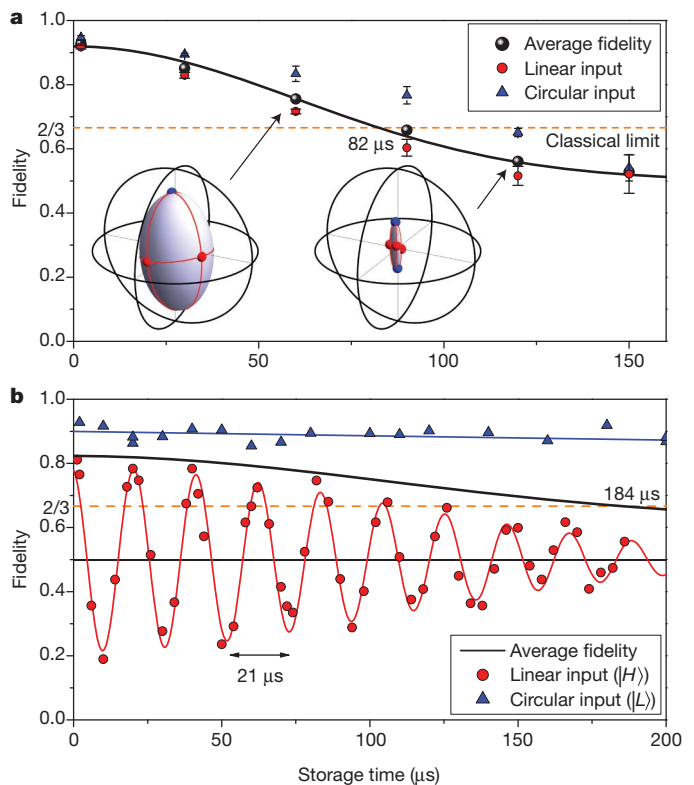


Figure 4 | Storage time. **a**, The fidelity between ideal input and measured output state as a function of storage time. A significantly slower decrease for the circular polarizations (blue triangles) compared to all linear polarizations (red circles) can be observed. The error bars denote the scatter for individual polarizations. For the average fidelity (black dots and black Gaussian fit), the classical limit of $2/3$ is reached after 82 μs . For the retrieved photon, the evolution of the Poincaré sphere in Stokes space confirms that the main decoherence mechanism is dephasing between the Zeeman states that encode $|R\rangle$ and $|L\rangle$. **b**, Increased storage times can be obtained by applying a constant guiding field of $B_{\text{guide}} = 34$ mG along the cavity axis. The resulting Larmor precession of the atomic Zeeman coherence results in an oscillation of the retrieved polarization for linearly polarized states (only $|H\rangle$ is shown) with a period of 21 μs . The net effect is a stabilization of the coherence of the atomic qubit and an increase in the average achievable storage time by a factor of more than 2. In **b** the average fidelity is defined as the mean of the fidelities for $|H\rangle$ (doubly weighted) and $|L\rangle$.

atomic $F = 1$ state, its use as the bright state in hyperfine state detection¹⁵ will herald the storage without degradation of the qubit¹⁴. Such a heralded memory can be used to perform a quantum non-demolition measurement on the presence of a photonic qubit. Furthermore, the demonstrated mapping of a photon state onto an atom can be used to realize a deterministic entanglement scheme between remote atomic qubits: if a first emitting atom is entangled with a photon¹¹, the absorption of this photon in a memory atom directly produces entanglement between the two atoms⁸. This establishes single-atom cavity systems as universal nodes of a future quantum network.

Received 20 January; accepted 9 March 2011.

Published online 1 May 2011.

1. Lvovsky, A. I., Sanders, B. C. & Tittel, W. Optical quantum memory. *Nature Photon.* **3**, 706–714 (2009).
2. Matsukevich, D. N. *et al.* Entanglement of remote atomic qubits. *Phys. Rev. Lett.* **96**, 030405 (2006).
3. Choi, K. S., Deng, H., Laurat, J. & Kimble, H. J. Mapping photonic entanglement into and out of a quantum memory. *Nature* **452**, 67–71 (2008).
4. Tanji, H., Ghosh, S., Simon, J., Bloom, B. & Vuletić, V. Heralded single-magnon quantum memory for photon polarization states. *Phys. Rev. Lett.* **103**, 043601 (2009).
5. Jin, X.-M. *et al.* Quantum interface between frequency-uncorrelated down-converted entanglement and atomic-ensemble quantum memory. Preprint at (<http://arxiv.org/abs/1004.4691>) (2010).

6. Saglamyurek, E. *et al.* Broadband waveguide quantum memory for entangled photons. *Nature* **469**, 512–515 (2011).
7. Clausen, C. *et al.* Quantum storage of photonic entanglement in a crystal. *Nature* **469**, 508–511 (2011).
8. Cirac, J. I., Zoller, P., Kimble, H. J. & Mabuchi, H. Quantum state transfer and entanglement distribution among distant nodes in a quantum network. *Phys. Rev. Lett.* **78**, 3221–3224 (1997).
9. Blinov, B. B., Moehring, D. L., Duan, L.-M. & Monroe, C. Observation of entanglement between a single trapped atom and a single photon. *Nature* **428**, 153–157 (2004).
10. Volz, J. *et al.* Observation of entanglement of a single photon with a trapped atom. *Phys. Rev. Lett.* **96**, 030404 (2006).
11. Wilk, T., Webster, S. C., Kuhn, A. & Rempe, G. Single-atom single-photon quantum interface. *Science* **317**, 488–490 (2007).
12. Boozer, A. D., Boca, A., Miller, R., Northup, T. E. & Kimble, H. J. Reversible state transfer between light and a single trapped atom. *Phys. Rev. Lett.* **98**, 193601 (2007).
13. Togan, E. *et al.* Quantum entanglement between an optical photon and a solid-state spin qubit. *Nature* **466**, 730–734 (2010).
14. Lloyd, S., Shahriar, M. S., Shapiro, J. H. & Hemmer, P. R. Long distance, unconditional teleportation of atomic states via complete Bell state measurements. *Phys. Rev. Lett.* **87**, 167903 (2001).
15. Bochmann, J. *et al.* Lossless state detection of single neutral atoms. *Phys. Rev. Lett.* **104**, 203601 (2010).
16. Piro, N. *et al.* Heralded single-photon absorption by a single atom. *Nature Phys.* **7**, 17–20 (2011).
17. Jaksch, D. *et al.* Fast quantum gates for neutral atoms. *Phys. Rev. Lett.* **85**, 2208–2211 (2000).
18. Briegel, H.-J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
19. Mücke, M. *et al.* Electromagnetically induced transparency with single atoms in a cavity. *Nature* **465**, 755–758 (2010).
20. Kampschulte, T. *et al.* Optical control of the refractive index of a single atom. *Phys. Rev. Lett.* **105**, 153603 (2010).
21. Hennrich, M., Legero, T., Kuhn, A. & Rempe, G. Vacuum-stimulated Raman scattering based on adiabatic passage in a high-finesse optical cavity. *Phys. Rev. Lett.* **85**, 4872–4875 (2000).
22. Keller, M., Lange, B., Hayasaka, K., Lange, W. & Walther, H. Continuous generation of single photons with controlled waveform in an ion-trap cavity system. *Nature* **431**, 1075–1078 (2004).
23. Vasilev, G. S., Ljunggren, D. & Kuhn, A. Single photons made-to-measure. *N. J. Phys.* **12**, 063024 (2010).
24. Gorshkov, A. V., André, A., Lukin, M. D. & Sørensen, A. S. Photon storage in Λ -type optically dense atomic media. I. Cavity model. *Phys. Rev. A* **76**, 033804 (2007).
25. James, D. F. V., Kwiat, P. G., Munro, W. J. & White, A. G. Measurement of qubits. *Phys. Rev. A* **64**, 052312 (2001).
26. Bowdrey, M. D., Oi, D. K. L., Short, A., Banaszek, K. & Jones, J. Fidelity of single qubit maps. *Phys. Rev. Lett.* **294**, 258–260 (2002).
27. Curty, M. & Lütkenhaus, N. Intercept-resend attacks in the Bennett-Brassard 1984 quantum-key-distribution protocol with weak coherent pulses. *Phys. Rev. A* **71**, 062301 (2005).
28. Massar, S. & Popescu, S. Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.* **74**, 1259–1263 (1995).
29. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
30. Zhao, B. *et al.* A millisecond quantum memory for scalable quantum networks. *Nature Phys.* **5**, 95–99 (2009).
31. Zhao, R. *et al.* Long-lived quantum memory. *Nature Phys.* **5**, 100–104 (2009).
32. Radnaev, A. G. *et al.* A quantum memory with telecom-wavelength conversion. *Nature Phys.* **6**, 894–899 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Kiesel for discussions and A. Neuzner for experimental assistance. This work was supported by the Deutsche Forschungsgemeinschaft (Research Unit 635), by the European Union (Collaborative Project AQUITE) and by the Bundesministerium für Bildung und Forschung via IKT 2020 (QK_QuOReP). E.F. acknowledges support from the Alexander von Humboldt Foundation.

Author Contributions All authors contributed to the experiment, the analysis of the results and the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.R. (stephan.ritter@mpq.mpg.de).

Quantum annealing with manufactured spins

M. W. Johnson¹, M. H. S. Amin¹, S. Gildert¹, T. Lanting¹, F. Hamze¹, N. Dickson¹, R. Harris¹, A. J. Berkley¹, J. Johansson², P. Bunyk¹, E. M. Chapple¹, C. Enderud¹, J. P. Hilton¹, K. Karimi¹, E. Ladizinsky¹, N. Ladizinsky¹, T. Oh¹, I. Perminov¹, C. Rich¹, M. C. Thom¹, E. Tolkacheva¹, C. J. S. Truncik³, S. Uchaikin¹, J. Wang¹, B. Wilson¹ & G. Rose¹

Many interesting but practically intractable problems can be reduced to that of finding the ground state of a system of interacting spins; however, finding such a ground state remains computationally difficult¹. It is believed that the ground state of some naturally occurring spin systems can be effectively attained through a process called quantum annealing^{2,3}. If it could be harnessed, quantum annealing might improve on known methods for solving certain types of problem^{4,5}. However, physical investigation of quantum annealing has been largely confined to microscopic spins in condensed-matter systems^{6–12}. Here we use quantum annealing to find the ground state of an artificial Ising spin system comprising an array of eight superconducting flux quantum bits with programmable spin–spin couplings. We observe a clear signature of quantum annealing, distinguishable from classical thermal annealing through the temperature dependence of the time at which the system dynamics freezes. Our implementation can be configured *in situ* to realize a wide variety of different spin networks, each of which can be monitored as it moves towards a low-energy configuration^{13,14}. This programmable artificial spin network bridges the gap between the theoretical study of ideal isolated spin networks and the experimental investigation of bulk magnetic samples. Moreover, with an increased number of spins, such a system may provide a practical physical means to implement a quantum algorithm, possibly allowing more-effective approaches to solving certain classes of hard combinatorial optimization problems.

Physically interesting in their own right, systems of interacting spins also have practical importance for quantum computation¹⁵. One widely studied example is the Ising spin model, where spins may take on one of two possible values: up or down along a preferred axis. Many seemingly unrelated yet important hard problems, in fields ranging from artificial intelligence¹⁶ to zoology¹⁷, can be reformulated as the problem of finding the lowest energy configuration, or ground state, of an Ising spin system.

Quantum annealing has been proposed as an effective way for finding such a ground state^{2–5}. To implement a processor that uses quantum annealing to help solve difficult problems, we would need a programmable quantum spin system in which we could control individual spins and their couplings, perform quantum annealing and then determine the state of each spin. Until recently, physical investigation of quantum annealing has been confined to configurations achievable in condensed-matter systems, such as molecular nanomagnets^{6–10} or bulk solids with quantum critical behaviour^{11,12}. Unfortunately, these systems cannot be controlled or measured at the level of individual spins, and are typically investigated through the measurement of bulk properties. They are not programmable. Nuclear magnetic resonance techniques have been used to demonstrate a quantum annealing algorithm on three quantum spins¹⁸. Recently, three trapped ions were used to perform a quantum simulation of a small, frustrated Ising spin system¹⁹.

One possible implementation of an artificial Ising spin system involves superconducting flux quantum bits^{20–28} (qubits). We have

implemented such a spin system, interconnected as a bipartite graph, using an *in situ* reconfigurable array of coupled superconducting flux qubits¹⁴. The device fabrication is discussed in Methods and in Supplementary Information. The simplified schematic in Fig. 1a shows two superconducting loops in the qubit, each subject to an external flux bias Φ_{1x} or Φ_{2x} , respectively. The device dynamics can be modelled as a quantum mechanical double-well potential with respect to the flux, Φ_1 , in loop 1 (Fig. 1b). The barrier height, δU , is controlled by Φ_{2x} . The energy difference between the two minima, $2h$, is controlled by Φ_{1x} . The two lowest energy states of the system, corresponding to clockwise or anticlockwise circulating current in loop 1, are labelled $|\downarrow\rangle$ and $|\uparrow\rangle$, with flux localized in the left- or the right-hand well (Fig. 1b), respectively. If we consider only these two states (a valid restriction at low temperature), the qubit dynamics is equivalent to those of an Ising spin, and we treat the qubits as such in what follows. Qubits (spins) are

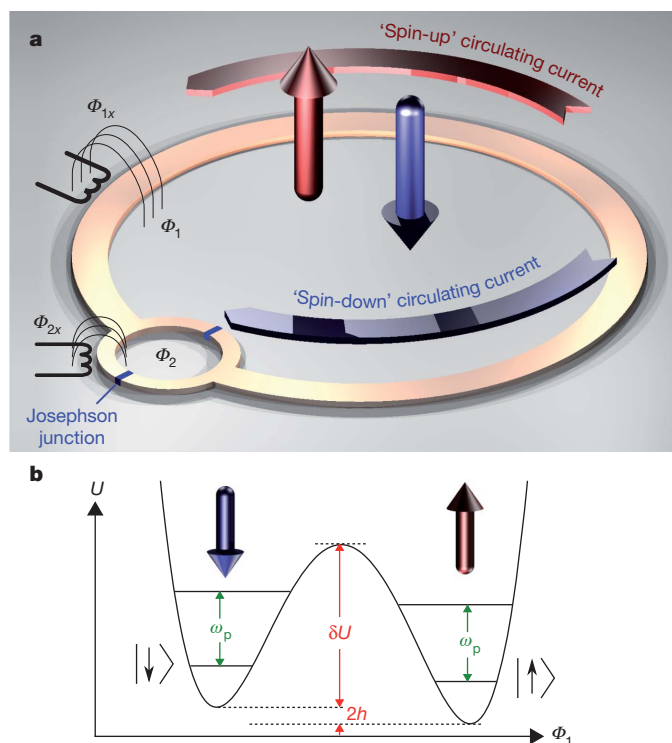


Figure 1 | Superconducting flux qubit. **a**, Simplified schematic of a superconducting flux qubit acting as a quantum mechanical spin. Circulating current in the qubit loop gives rise to a flux inside, encoding two distinct spin states that can exist in a superposition. **b**, Double-well potential energy diagram and the lowest quantum energy levels corresponding to the qubit. States $|\uparrow\rangle$ and $|\downarrow\rangle$ are the lowest two levels, respectively. The intra-well energy spacing is ω_p . The measurement detects magnetization, and does not distinguish between, say, $|\uparrow\rangle$ and excited states within the right-hand well. In practice, these excitations are exceedingly improbable at the time the state is measured.

¹D-Wave Systems Inc., 100-4401 Still Creek Drive, Burnaby, British Columbia V5C 6G9, Canada. ²Department of Natural Sciences, University of Agder, Post Box 422, NO-4604 Kristiansand, Norway.

³Department of Physics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada.

coupled together using programmable coupling elements²⁹ which provide a spin–spin coupling energy that is continuously tunable between ferromagnetic and antiferromagnetic coupling. This allows spins to favour alignment or anti-alignment, respectively.

The behaviour of this system can be described with an Ising model Hamiltonian

$$\mathcal{H}_P = \sum_{i=1}^N h_i \sigma_i^z + \sum_{i,j=1}^N J_{ij} \sigma_i^z \sigma_j^z \quad (1)$$

where for spin i σ_i^z is the Pauli spin matrix with eigenvectors $\{|\uparrow\rangle, |\downarrow\rangle\}$ and $2h_i$ is the energy bias; and $2J_{ij}$ is the coupling energy between the spins i and j . Our implementation allows each J_{ij} and h_i to be programmed independently within the constraints of the connectivity of our devices.

The quantum mechanical properties of the individual devices have been well characterized¹³, but we are interested in what happens when several of them are coupled together. It is reasonable to ask whether this manufactured, macroscopic (~ 1 mm) system of artificial spins behaves quantum mechanically. We report here on an experiment that demonstrates a signature of quantum annealing in a coupled set of eight artificial Ising spins.

Whereas thermal annealing uses progressively weaker thermal fluctuations to allow a system to explore its energy landscape and arrive at a low-energy configuration, quantum annealing uses progressively weaker quantum fluctuations, mediated by tunnelling. In both thermal and quantum annealing, a system starts with a mixture of all possible states: a classical mixed state in the former and a coherent superposition in the latter.

Quantum annealing can be performed by slowly changing the system Hamiltonian

$$\mathcal{H}(t) = \Gamma(t) \sum_{i=1}^N A_i \sigma_i^x + \Lambda(t) \mathcal{H}_P$$

where Γ decreases from one to zero and Λ increases from zero to one monotonically with time, and A_i parameterizes quantum mechanical tunnelling between $|\uparrow\rangle$ and $|\downarrow\rangle$.

At the beginning of the annealing, $\Gamma = 1$, $\Lambda = 0$ and the system is fully characterized by the transverse terms, $\sum_{i=1}^N A_i \sigma_i^x$. The ground state of this is a superposition of all states in the σ_z basis. It is straightforward to initialize the system in this state. During quantum annealing, the transverse term is gradually turned off ($\Gamma \rightarrow 0$) and the weight of the Ising Hamiltonian, \mathcal{H}_P , is increased ($\Lambda \rightarrow 1$) (Fig. 2b). If this annealing is done slowly enough, the system should remain in the ground state at all times, thus ending up in the ground state of \mathcal{H}_P (ref. 4).

The above description of quantum annealing is in the language of an ideal Ising spin system. Let us look more closely at what this means for an individual flux qubit. During annealing, the energy barrier, $\delta U(t)$, between the two wells is gradually raised (Fig. 2a). If thermal fluctuations are dominant, then the qubit dynamics may be viewed as thermal activation over the barrier with a rate that is proportional to $e^{-\delta U/k_B T}$ at a temperature T (k_B , Boltzmann's constant). This suggests that the dynamics stops when $\delta U \gg k_B T$. Because δU is increasing with time, this freezing out happens at $t \approx t_{\text{freeze}}^{\text{TA}}$, where $\delta U(t_{\text{freeze}}^{\text{TA}}) \approx k_B T$. Within the relevant regime, δU is nearly linear in time, therefore classically we expect $t_{\text{freeze}}^{\text{TA}}$ to be linearly dependent on T .

If, however, the dominant fluctuations are quantum mechanical, then the qubit may tunnel between the two wells, that is, between states $|\downarrow\rangle$ and $|\uparrow\rangle$. Raising the barrier, by increasing δU , reduces this tunnelling until at some point it becomes negligible. In this picture, we expect to find a quantum freeze-out time, $t_{\text{freeze}}^{\text{QA}}$, that is independent of (or at least very weakly dependent on) T . By measuring the T dependence of t_{freeze} , the time at which the system can no longer respond to changes in its energy landscape, we can determine whether classical thermal

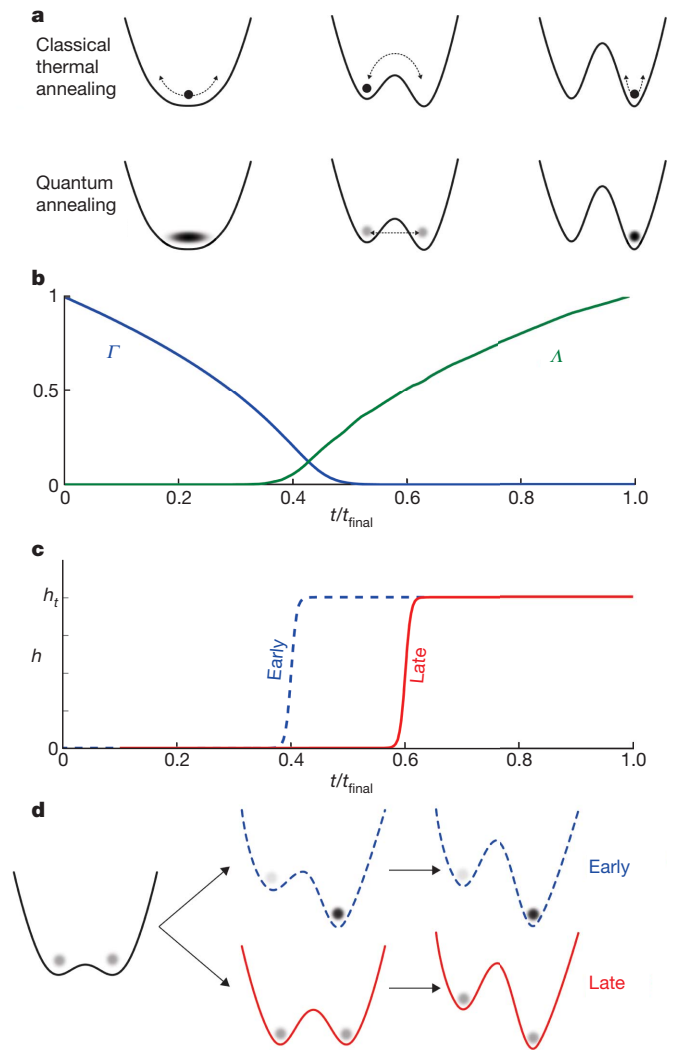


Figure 2 | Quantum annealing. **a**, Annealing is performed by gradually raising the energy barrier between states. In thermal annealing, when the barrier becomes much larger than $k_B T$ thermal excitation over the barrier eventually ceases, at some time $t_{\text{freeze}}^{\text{TA}}$. In quantum annealing, tunnelling between states also will eventually cease, at a time $t_{\text{freeze}}^{\text{QA}}$. **b**, The value of the parameters Γ and Λ during annealing are not independent of each other in the flux qubit. The annealing ends at $t_{\text{final}} = 148 \mu\text{s}$. **c**, Changing the value of $h(t)$ (see **d**) at various points during annealing can be used to probe the freeze-out time, t_{freeze} . **d**, Double-well potential during annealing. If h is turned on early enough (blue line), the system follows the ground state through annealing and reaches the final ground state of equation (1) with high probability. If $h(t)$ is turned on too late (red line), the state probabilities are determined by the earlier Hamiltonian, for which $h = 0$.

activation or quantum tunnelling is the dominant effect governing qubit dynamics.

Here we modify this annealing procedure to perform a specialized experiment that permits us to distinguish between these two cases, by allowing the h_i in equation (1) to be time dependent. We measure the ‘step response’ of the system to rapid changes in h (rapid by comparison with changes in Γ and Λ) at different stages during the annealing process. In this way, we are able to measure t_{freeze} . By measuring t_{freeze} as a function of T , we can infer whether the system dynamics is dominated by thermal or quantum fluctuations.

We abruptly increase h from zero to a level h_t at a delay time t_d during annealing as shown in Fig. 2c, and then measure the probability of the spin being in either configuration at the end of annealing. If h is switched on very early in the annealing process, while the barrier δU is still small in comparison with the thermal or quantum transition

energy scales, then the qubit will quickly respond and will be able to evolve into the lower energy well, such that P_{\uparrow} , the probability of the spin being in state $|\uparrow\rangle$, is greater than $1/2$. The value of P_{\uparrow} will depend on both h_t and T , as the system will strive to achieve a Boltzmann distribution of its population statistics between $|\downarrow\rangle$ and $|\uparrow\rangle$. However, if h is not turned on until after the barrier has been raised sufficiently high ($t_d > t_{\text{freeze}}$), the system will not be able to follow it and will be equally likely to settle into either potential energy well, such that $P_{\uparrow} \approx 1/2$. These two situations are illustrated in Fig. 2d. For intermediate values of t_d , the qubit will only partly succeed in responding to the sudden application of the bias h .

Example plots of measured P_{\uparrow} values versus t_d for a single qubit at different temperatures are shown in Fig. 3a. In this case, $h_t = 2.55 \pm 0.04$ GHz and $\Delta = 9.0 \pm 0.2$ GHz. Experimental parameters controlling the annealing process are discussed in Supplementary Information. As expected, P_{\uparrow} shows an initial ($t_d \approx 0$) T dependence and then converges to $1/2$ at late delay times. These curves were numerically fitted to extract t_{freeze} , the time at the middle of this transition region, which in turn is plotted versus T in Fig. 3b. The curve used for the fitting was obtained by numerical simulations of this

process using a quantum mechanical model as discussed in the Supplementary Information.

In addition to the experimental results, in Fig. 3b we show the results of three different numerical simulations. In all three cases, the model parameters were independently measured for the individual devices, leaving no free parameters. A simulation, based on a classical model, treated the flux in the two superconducting loops as the coordinates of a discrete particle inside the two-dimensional flux qubit potential, and then coupled that particle to a thermal bath. The dynamics was simulated by numerically solving the Langevin equation, as described in Supplementary Information. The classically simulated t_{freeze} value varies linearly with T , as expected. The other two simulations involved solving a quantum mechanical model of a flux qubit coupled to a thermal bath in which only the two or, respectively, four lowest-lying energy levels of the flux qubit were kept. The dynamics was simulated by numerically solving a non-Markovian density matrix equation of motion (Supplementary Information). These two models will be referred to here as the two-level and four-level quantum models.

The experimental results clearly show a saturation of t_{freeze} below 45 mK, in agreement with both the two-level and the four-level quantum models and in disagreement with the classical model. The experimental data deviate from the two-level model above 45 mK, as the upper energy levels in the flux qubit start to become thermally occupied. The four-level quantum model describes the behaviour of the system well up to 80 mK, where more energy levels start to be occupied. The experimental data asymptotically approach the classical simulation results at higher temperatures. We propose that if the quantum mechanical modelling were extended by keeping even more energy levels, then it would reproduce the data to ever higher temperatures.

Both the measured and the simulated (four-level quantum model) T dependence of P_{\uparrow} for $t_d \approx 0$ are shown in the inset of Fig. 3a. Because this probability has a strong T dependence for $t_d < t_{\text{freeze}}$, its measurement provides us with an independent check on the effective temperature of the spin system in this regime. Moreover, because the probability does not saturate at 45 mK, where t_{freeze} saturates, it is a clear indication that saturation of t_{freeze} is not a result of saturation of qubit temperature. The key conclusion we draw from Fig. 3b is that our qubit dynamics is best characterized as being quantum mechanical in nature for $T \lesssim 80$ mK. The system evolves to its ground state through a process of quantum annealing. But so far we have shown this only for an individual qubit. It remains to be shown whether quantum annealing can be performed on several spins coupled together.

To investigate this, we now configure our array into a chain of eight ferromagnetically coupled artificial spins (Fig. 4), with $J_{i,i+1} = -J$ for $i = 1, 2, \dots, 7$ along the chain and $J_{ij} = 0$ otherwise. In our experiment, we used $J = 12.78$ GHz, which is near the maximum available for the couplers. The lowest-energy configurations of this system correspond to the two ferromagnetic states $|\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\rangle$ and $|\downarrow\downarrow\downarrow\downarrow\downarrow\downarrow\rangle$. Applying strong but opposing biases, $h_B = \pm 2J$, to the ends of the chain introduces frustration into the system, and the lowest-energy configuration will have a break in the ferromagnetic order; this is known as a domain wall (where the spins change direction). For example, we depict the state $|\uparrow\uparrow\uparrow\downarrow\downarrow\downarrow\rangle$ in Fig. 4a, for which the domain wall is the middle of the chain.

In our step response experiment with the spins configured as a chain, all six spins internal to the chain begin annealing with $h = 0$. In this situation, it is energetically equivalent for the domain wall to be between any adjacent pair of spins, and each such state should occur with probability $P = 1/7$. If we leave $h = 0$ for too long ($t_d \gg t_{\text{freeze}}$), we expect to observe this distribution of single-domain-wall states. At $t = t_d$, we apply a uniform bias, $h_t = 0.1J$, to the six intermediate spins. Now the ground state is $|\uparrow\uparrow\uparrow\uparrow\uparrow\downarrow\rangle$, with the domain wall at the right-hand end of Fig. 4b. More system energy is required for the domain wall to occupy positions to the left in Fig. 4. If $t_d \ll t_{\text{freeze}}$, we should observe state $|\uparrow\uparrow\uparrow\uparrow\uparrow\downarrow\rangle$ occurring with probability $P > 1/7$. As in the single-qubit case, we measure t_{freeze} by finding the transition point

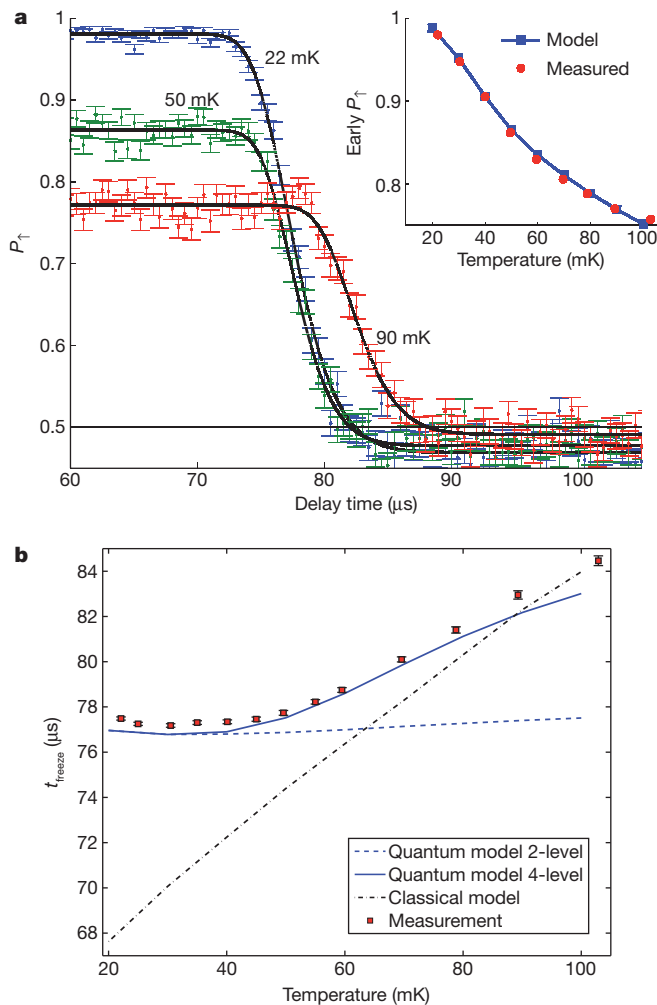


Figure 3 | Single-qubit results. **a**, Measured final ground-state probability, P_{\uparrow} , in a single qubit versus the delay time, t_d , of a step $h_t = 2.55 \pm 0.04$ GHz in energy bias, for $T = 22$ mK (blue), 50 mK (green) and 90 mK (red). The solid lines are the result of fits used to extract the freeze-out time, t_{freeze} . Inset, measured and simulated (four-level quantum model) T dependence of P_{\uparrow} for $t_d \approx 0$. **b**, Measured t_{freeze} versus T (red points). We also show simulated plots of t_{freeze} from two-level (dashed blue) and four-level (solid blue) quantum mechanical models and from a classical model of the qubit (black). Error bars, 1σ s.e.

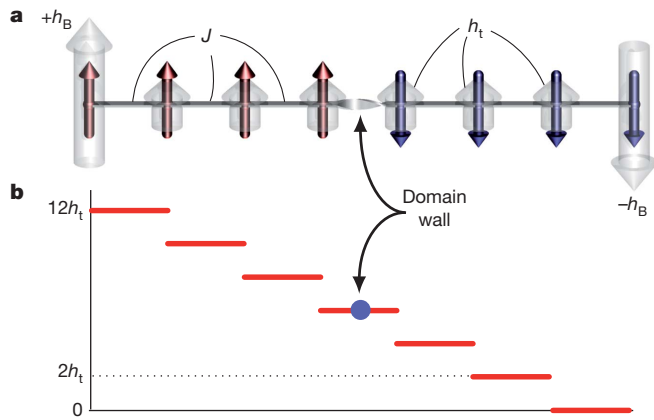


Figure 4 | Eight-qubit ferromagnetic chain. **a**, Chain of eight ferromagnetically coupled qubits with uniform coupling coefficient $J_{i,i+1} = -J < 0$ for $i = 1, 2, \dots, 7$. The two end qubits are biased in opposite directions with $h_B = \pm 2J$, such that a domain wall has to form within the chain. All middle qubits are biased with a target $h_t = 0.1J$. The configuration depicted is an excited state. The faint grey arrows indicate the spin biases h_i . **b**, Effective energy of the spin state corresponding to there being a domain wall at each position along the chain at finite h_t . The ground state is the rightmost site.

between these two state distributions (Fig. 5a). As before, we are able to determine the dominant mechanism (thermal or quantum annealing) by measuring the T dependence of t_{freeze} .

A summary of the experimental results for the eight-qubit chain is shown in Fig. 5b. As with the single-qubit case, the experimentally determined t_{freeze} values show saturation at low T and a crossover to near-linear T dependence for $T \gtrsim 45$ mK. In this case, the classical model treats the fluxes of all eight qubits as coordinates of a discrete particle in a sixteen-dimensional potential. The classical model does not capture the behaviour observed at low T . However, the quantum models quantitatively agree with the experimental results for $T \lesssim 50$ mK. At higher temperatures, the classical model and the four-level quantum model are both in qualitative agreement with the experimental results.

The saturation of t_{freeze} at low T for the single-spin and eight-spin systems is a clear signature of quantum annealing. It cannot be explained by an experimental failure to reach lower T , as $P_{\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow}$ for $t_d \ll t_{\text{freeze}}$, follows its expected temperature dependence at low T (Fig. 5a, inset). Nor can it be explained by classical thermal activation processes, because for these lowering T would always decrease the rate of thermal activation. This means that, classically, freeze-out should happen earlier in the evolution, where the barrier is smaller, that is, saturation is not possible. This qualitative argument is independent of the detailed model used to describe classical dynamics. The low-temperature behaviour of t_{freeze} in this system of eight coupled artificial spins cannot be explained by thermal activation but is naturally explained by quantum tunnelling. This measurement and its result are reminiscent of the T -dependent escape rate measurements in the pioneering works on macroscopic quantum tunnelling^{30,31}, which demonstrated a clear signature of quantum tunnelling in current-biased Josephson junctions.

This brings us to our main conclusion: a programmable artificial spin system manufactured as an integrated circuit can be used to implement a quantum algorithm. The experiments presented here constitute a step between understanding single-qubit annealing and understanding the multi-qubit processes that could be used to find low-energy configurations in a realistic adiabatic quantum processor. In addition to its problem-solving potential, a system such as this also provides an interesting test bed for investigating the physics of interacting quantum spins, and is an important step in an ongoing investigation into much more complex spin systems realized using this type of architecture. Although our manufactured spin system is not yet a

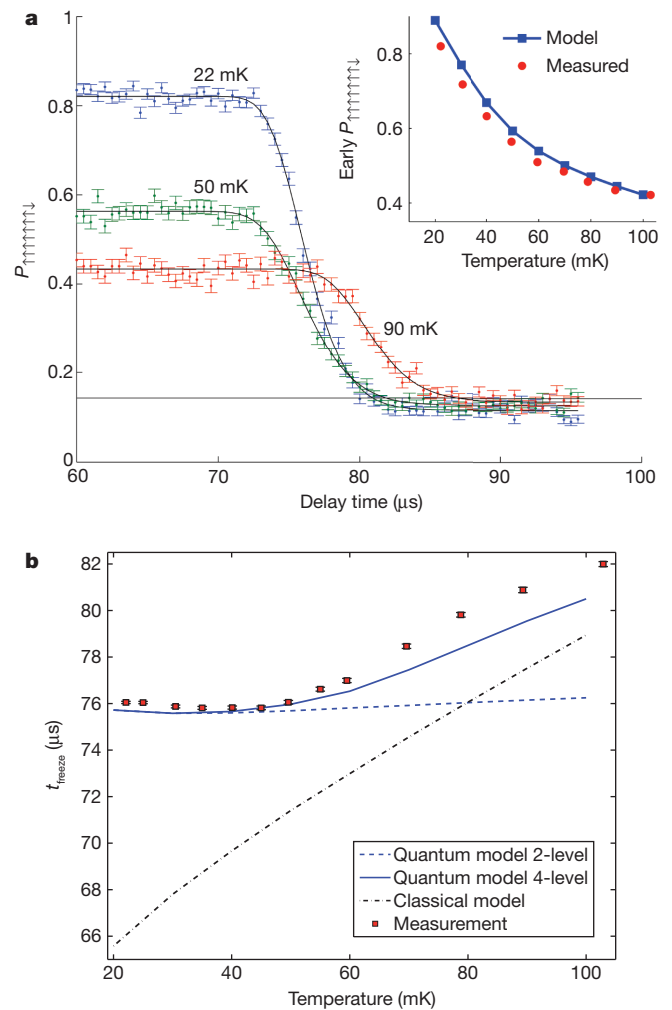


Figure 5 | Results for the eight-qubit ferromagnetic chain. **a**, Measured final ground-state probability, $P_{\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow}$, in the eight-qubit chain versus t_d for $h_t = 0.1J$ and $T = 22$ mK (blue), 50 mK (green) and 90 mK (red). The solid lines are the result of fits used to extract the freeze-out time, t_{freeze} . Inset, measured and simulated (four-level quantum model) T dependence of $P_{\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow}$ for $t_d \approx 0$. **b**, Measured t_{freeze} versus T (red points). We also show simulated plots of t_{freeze} from two-level (dashed blue) and four-level (solid blue) quantum mechanical models and from a classical model of the qubits (black). Error bars, 1σ s.e.

universal quantum computer¹⁵, by adding a new type of coupler between the qubits, universal quantum computation would become possible³².

METHODS SUMMARY

Sample fabrication. We fabricated samples in a four-niobium-layer superconducting integrated circuit process using a standard Nb/AlO_x/Nb trilayer, a TiPt resistor layer and planarized SiO₂ dielectric layers applied by plasma-enhanced chemical vapour deposition. Design rules included 0.25- μm lines and spaces for wiring layers and a minimum junction diameter of 0.6 μm . Circuit details are discussed in ref. 13.

Thermometry. We measured the effective device temperature attained during these experiments in two ways. The first is based on analysis of the single-qubit macroscopic resonant tunnelling. The second is based on measurement of P_1 versus $\Phi_{1,x}$ at equilibrium and at a fixed barrier height. Both methods are described in ref. 33 and in Supplementary Information. Both measurements generally agreed with the ruthenium oxide thermometer on the dilution refrigerator mixing chamber to within a few millikelvin over the range of temperatures used in the experiment.

Confirmation of the thermometry comes from the agreement between the measured T dependence of $P_1(t_d \ll t_{\text{freeze}})$ and that predicted by the four-level quantum model (insets of Fig. 3a and Fig. 5a). This is discussed further in Supplementary Information.

Annealing. Annealing was performed by sweeping Φ_{2x} (Fig. 1b) from $0.592\Phi_0$ to $0.652\Phi_0$ linearly over a period of 148 μs , where Φ_0 is the magnetic flux quantum. These values bracket the point at which the qubit becomes bistable. The devices used are those analysed in ref. 14. (See Supplementary Information for more details.)

Received 30 June 2010; accepted 15 March 2011.

- Barahona, F. On the computational complexity of Ising spin glass models. *J. Phys. Math. Gen.* **15**, 3241–3253 (1982).
- Kadowaki, T. & Nishimori, H. Quantum annealing in the transverse Ising model. *Phys. Rev. E* **58**, 5355–5363 (1998).
- Finnila, A. B., Gomez, M. A., Sebenik, C., Stenson, C. & Doll, J. D. Quantum annealing: a new method for minimizing multidimensional functions. *Chem. Phys. Lett.* **219**, 343–348 (1994).
- Farhi, E. *et al.* A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem. *Science* **292**, 472–475 (2001).
- Hogg, T. Quantum search heuristics. *Phys. Rev. A* **61**, 052311 (2000).
- Wernsdorfer, W. Molecular nanomagnets: towards molecular spintronics. *Int. J. Nanotechnol.* **7**, 497–522 (2010).
- Carretta, S., Liviotti, E., Magnani, N., Santini, P. & Amoretti, G. S mixing and quantum tunneling of the magnetization in molecular nanomagnets. *Phys. Rev. Lett.* **92**, 207205 (2004).
- Caciuffo, R. *et al.* Spin dynamics of heterometallic Cr_7M wheels ($\text{M} = \text{Mn, Zn, Ni}$) probed by inelastic neutron scattering. *Phys. Rev. B* **71**, 174407 (2005).
- Guidi, T. *et al.* Inelastic neutron scattering study of the molecular grid nanomagnet $\text{Mn}[\text{3} \times \text{3}]$. *Phys. Rev. B* **69**, 104432 (2004).
- Waldmann, O., Guidi, T., Carretta, S., Mondelli, C. & Dearden, A. L. Elementary excitations in the cyclic molecular nanomagnet Cr_8 . *Phys. Rev. Lett.* **91**, 237202 (2003).
- Brooke, J., Bitko, D., Rosenbaum, T. F. & Aeppli, G. Quantum annealing of a disordered magnet. *Science* **284**, 779–781 (1999).
- Ghosh, S. & Rosenbaum, T. F. Aeppli, G. & Coppersmith, S. N. Entangled quantum state of magnetic dipoles. *Nature* **425**, 48–51 (2003).
- Harris, R. *et al.* Experimental demonstration of a robust and scalable flux qubit. *Phys. Rev. B* **81**, 134510 (2010).
- Harris, R. *et al.* Experimental investigation of an eight-qubit unit cell in a superconducting optimization processor. *Phys. Rev. B* **82**, 024511 (2010).
- Aharonov, D. *et al.* Adiabatic quantum computation is equivalent to standard quantum computation. *SIAM J. Comput.* **37**, 166–194 (2007).
- Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504–507 (2006).
- Chen, X. & Tompa, M. Comparative assessment of methods for aligning multiple genome sequences. *Nature Biotechnol.* **28**, 567–572 (2010).
- Steffen, M., van Dam, W., Hogg, T., Breyta, G. & Chuang, I. Experimental implementation of an adiabatic quantum optimization algorithm. *Phys. Rev. Lett.* **90**, 067903 (2003).
- Kim, K. *et al.* Quantum simulation of frustrated Ising spins with trapped ions. *Nature* **465**, 590–593 (2010).
- Lupaşcu, A. *et al.* Quantum non-demolition measurement of a superconducting two-level system. *Nature Phys.* **3**, 119–125 (2007).
- Berns, D. M. *et al.* Amplitude spectroscopy of a solid-state artificial atom. *Nature* **455**, 51–58 (2008).
- Poletto, S. *et al.* Coherent oscillations in a superconducting tunable flux qubit manipulated without microwaves. *N. J. Phys.* **11**, 013009 (2009).
- DiCarlo, L. *et al.* Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature* **460**, 240–244 (2009).
- Bennett, D. A. *et al.* Decoherence in rf SQUID qubits. *Quantum Inf. Process.* **8**, 217–243 (2009).
- Yoshihara, F., Nakamura, Y. & Tsai, J. S. Correlated flux noise and decoherence in two inductively coupled flux qubits. *Phys. Rev. B* **81**, 132502 (2010).
- Il'ichev, E. *et al.* Multiphoton excitations and inverse population in a system of two flux qubits. *Phys. Rev. B* **81**, 012506 (2010).
- Vion, D. *et al.* Manipulating the quantum state of an electrical circuit. *Science* **296**, 886–889 (2002).
- Burkard, G., Koch, R. H. & DiVincenzo, D. P. Multilevel quantum description of decoherence in superconducting qubits. *Phys. Rev. B* **69**, 064503 (2004).
- Harris, R. *et al.* Compound Josephson-junction coupler for flux qubits with minimal crosstalk. *Phys. Rev. B* **80**, 052506 (2009).
- Voss, R. F. & Webb, R. A. Macroscopic quantum tunneling in 1- μm Nb Josephson junctions. *Phys. Rev. Lett.* **47**, 265–268 (1981).
- Devoret, M. H., Martinis, J. M. & Clarke, J. Measurements of macroscopic quantum tunneling out of the zero-voltage state of a current-biased Josephson junction. *Phys. Rev. Lett.* **55**, 1908–1911 (1985).
- Biamonte, J. D. & Love, P. J. Realizable Hamiltonians for universal adiabatic quantum computers. *Phys. Rev. A* **78**, 012352 (2008).
- Harris, R. *et al.* Probing noise in flux qubits via macroscopic resonant tunneling. *Phys. Rev. Lett.* **101**, 117003 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank J. Preskill, A. Kitaev, D. A. Lidar, F. Wilhelm, A. Lupaşcu, A. Blais, T. A. Brun, P. Smith, F. Altomare, E. Hoskinson, T. Przybysz, T. Mahon and R. Neufeld for discussions. We are grateful to the volunteers of the AQUA@home BOINC project for their help in running the classical simulations.

Author Contributions M.H.S.A. and M.W.J. developed the idea for the experiment; M.W.J. conducted the experiment; T.L., R.H., M.W.J. and J.W. conducted supporting experiments; M.H.S.A. developed the theory; N.D., F.H. and M.H.S.A. developed simulation code; N.D., M.H.S.A., M.W.J., F.H. and C.J.S.T. performed simulations and analysed results; M.W.J., M.H.S.A., S.G. and R.H. wrote the article; M.W.J., S.G., M.H.S.A. and N.D. generated the figures; A.J.B., R.H., J.J., M.W.J., T.L., I.P., E.M.C. and B.W. developed measurement algorithms and testing software; C.R., S.U. and M.C.T. achieved the low-magnetic-field environment for the device; C.E. and C.R. mounted the sample, P.B., E.T., A.J.B., R.H., J.J., M.W.J. and T.L. designed the devices; E.L., N.L. and T.O. fabricated the devices; M.C.T. and S.U. developed the testing apparatus; K.K. allowed use of BOINC for classical simulations; J.P.H. and G.R. provided logistical support; and J.P.H. selected the chip.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.W.J. ([mwjohnson@dwavesys.com](mailto:mwjjohnson@dwavesys.com)).

Spin crossover and iron-rich silicate melt in the Earth's deep mantle

Ryuichi Nomura^{1,2}, Haruka Ozawa^{1,3}, Shigehiko Taten¹, Kei Hirose^{1,3}, John Hernlund⁴, Shunsuke Muto⁵, Hirofumi Ishii⁶ & Nozomu Hiraoka⁶

A melt has greater volume than a silicate solid of the same composition. But this difference diminishes at high pressure, and the possibility that a melt sufficiently enriched in the heavy element iron might then become more dense than solids at the pressures in the interior of the Earth (and other terrestrial bodies) has long been a source of considerable speculation^{1,2}. The occurrence of such dense silicate melts in the Earth's lowermost mantle would carry important consequences for its physical and chemical evolution and could provide a unifying model for explaining a variety of observed features in the core-mantle boundary region³. Recent theoretical calculations⁴ combined with estimates of iron partitioning between (Mg,Fe)SiO₃ perovskite and melt at shallower mantle conditions^{5–7} suggest that melt is more dense than solids at pressures in the Earth's deepest mantle, consistent with analysis of shockwave experiments⁸. Here we extend measurements of iron partitioning over the entire mantle pressure range, and find a precipitous change at pressures greater than ~76 GPa, resulting in strong iron enrichment in melts. Additional X-ray emission spectroscopy measurements on (Mg_{0.95}Fe_{0.05})SiO₃ glass indicate a spin collapse around 70 GPa, suggesting that the observed change in iron partitioning could be explained by a spin crossover of iron (from high-spin to low-spin) in silicate melt. These results imply that (Mg,Fe)SiO₃ liquid becomes more dense than coexisting solid at ~1,800 km depth in the lower mantle. Soon after the Earth's formation, the heat dissipated by accretion and internal differentiation could have produced a dense melt layer up to ~1,000 km in

thickness underneath the solid mantle. We also infer that (Mg,Fe)SiO₃ perovskite is on the liquidus at deep mantle conditions, and predict that fractional crystallization of dense magma would have evolved towards an iron-rich and silicon-poor composition, consistent with seismic inferences of structures in the core-mantle boundary region.

Our melting experiments were performed on samples with bulk composition (Mg_{0.89}Fe_{0.11})₂SiO₄ at pressures from 20 to 159 GPa in a laser-heated diamond-anvil cell (DAC; Supplementary Table 1). The heating duration was short in order to avoid anomalous thermal diffusion (Supplementary Information), but this prevented us from measuring the melting temperature. Nevertheless, the upper and lower bounds of the temperature in our experiment are given by the liquidus temperature of Mg₂SiO₄ and the solidus temperature of natural peridotite, respectively (see Methods and Supplementary Fig. 1). Samples were recovered from the DAC and examined with a high-resolution field-emission-type electron probe micro-analyser (FE-EPMA). Recovered specimens exhibited a concentric texture that reflected the temperature distribution during heating (Fig. 1), which is similar to that observed in conventional multi-anvil experiments^{5–7}. We consistently found a pocket at the hottest part of the sample that possessed non-stoichiometric composition, which we interpret as quenched partial melt. The (Mg + Fe)/Si molar ratio of this quenched melt increased with pressure, from 1.50 at 36 GPa to 2.56 at 159 GPa (Supplementary Fig. 2). The melt pocket was surrounded by a single-phase solid layer (ferropericlase or perovskite, depending on pressure), which we interpret to be

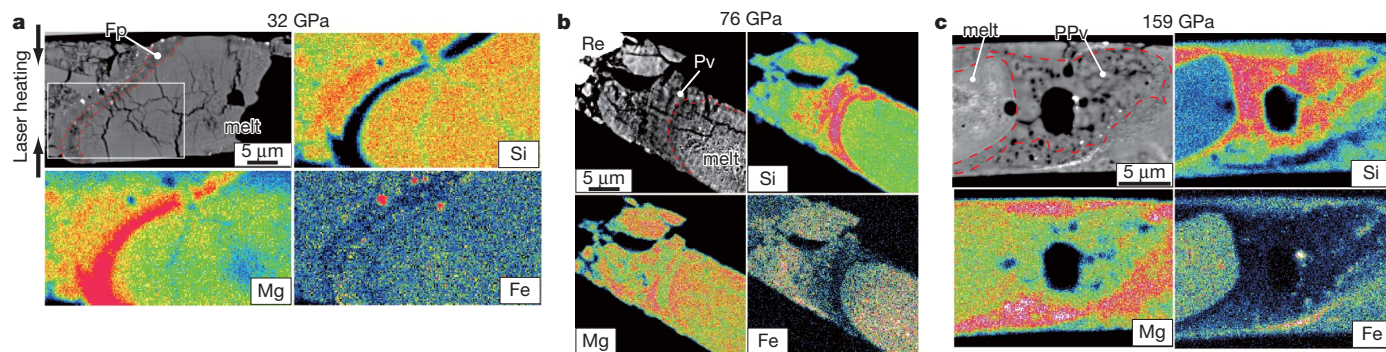


Figure 1 | Backscattered electron images and X-ray maps for Si, Mg and Fe for samples recovered from high-pressure melting experiments. **a**, At 32 GPa, when ferropericlase (Fp) is the liquidus phase; **b**, at 76 GPa, when perovskite (Pv) is the crystalline phase in contact with the quenched melt pocket; and **c**, at 159 GPa in the stability field of post-perovskite (PPv). Quenched melt was found at the centre of the sample, where the temperature

was highest. Metallic iron was observed at the edge of the laser-heated area in all samples, where a strong temperature gradient existed²⁸. It was also found in the melt pocket, but only above 36 GPa where the liquidus phase was perovskite or post-perovskite. Arrows in **a** represent the directions of the laser beams for heating. See Supplementary Information for the valence state of iron in the partial melt.

¹Department of Earth and Planetary Sciences, Tokyo Institute of Technology, Meguro, Tokyo 152-8551, Japan. ²Department of Earth and Planetary Sciences, University of Tokyo, Bunkyo, Tokyo 113-0033, Japan. ³Institute for Research on Earth Evolution, Japan Agency for Marine-Earth Science and Technology, Yokosuka, Kanagawa 237-0061, Japan. ⁴Department of Earth and Planetary Science, University of California, Berkeley, California 94720, USA. ⁵Department of Materials, Physics and Energy Engineering, Nagoya University, Chikusa, Nagoya 464-8603, Japan. ⁶National Synchrotron Radiation Research Center, Hsinchu 30076, Taiwan.

the phase that crystallizes first as temperature decreases (the liquidus phase)^{5,7}.

In this study, the liquidus phase was ferropericlase from 20 to 36 GPa and was subsequently replaced by perovskite at higher pressures (both were in contact with the melt pool at 36 GPa) (Fig. 1). Considering the difference in Mg/Si ratio of the samples⁷, this is consistent with observations⁵ of material with peridotite composition where the liquidus phase changes from ferropericlase to perovskite above 31 GPa. Although phase identification was not made, post-perovskite should have been formed in our experiments performed at 143 and 159 GPa. The change in liquidus phase from ferropericlase to perovskite at 36 GPa indicates that the eutectic melt composition becomes more Mg-rich at higher pressures. This is in agreement with the increase in (Mg + Fe)/Si molar ratio of partial melt with increasing pressure (Supplementary Fig. 2), although the partial melt composition also depends on the degree of melting.

The Fe–Mg distribution coefficient between perovskite/post-perovskite and melt, $K_D = ([Fe^{Pv}]/[Mg^{Pv}])/([Fe^{melt}]/[Mg^{melt}])$, was determined in the pressure range from 36 to 159 GPa (Fig. 2a), where perovskite/post-perovskite was the liquidus phase. Although the quenched melt

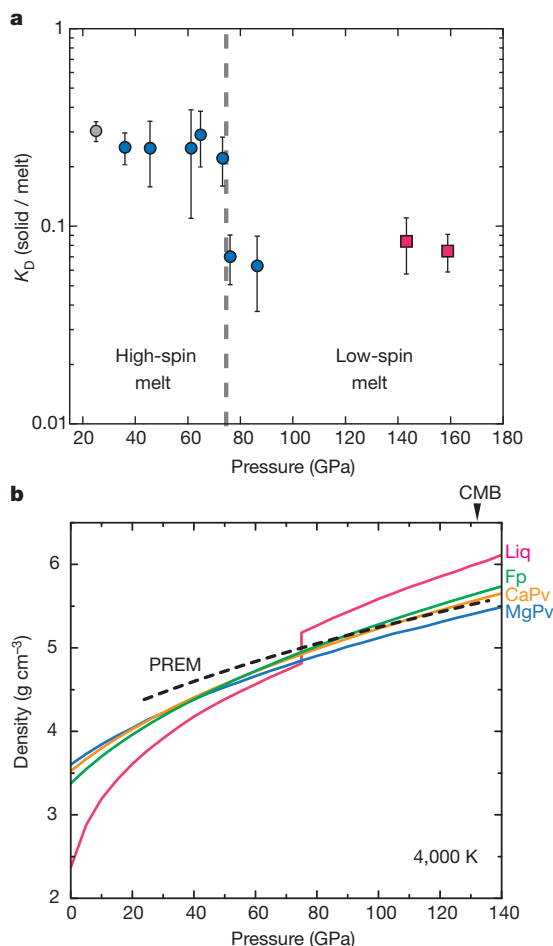


Figure 2 | Change in Fe–Mg distribution coefficient and calculated density profiles. **a**, $K_D = ([Fe^{Pv}]/[Mg^{Pv}])/([Fe^{melt}]/[Mg^{melt}])$ between perovskite (blue circles) or post-perovskite (red squares) and melt; the values drop sharply at pressures above 76 GPa, probably due to the effect of the spin crossover of iron in silicate melt (see Fig. 3). Previous experimental datum obtained at 25 GPa using a multi-anvil apparatus is shown by a grey circle⁶. Error bars were estimated from uncertainties (1σ) in both solid and liquid compositions. **b**, Density of the (Mg,Fe)SiO₃ liquid coexisting with (Mg_{0.92}Fe_{0.08})SiO₃ perovskite calculated for 4,000 K using the newly obtained Fe–Mg partitioning data. Data for (Mg_{0.86}Fe_{0.14})O ferropericlase²⁹, Ca-perovskite³⁰ and PREM¹⁹ are also shown for comparison. Liq, liquid; Fp, ferro-periclase; CaPv, calcium silicate perovskite; MgPv, magnesium silicate perovskite.

pocket contained multiple valence states of iron (Fig. 1b, c), we consider that all Fe was present as Fe²⁺ at high temperature (see Supplementary Information). The obtained K_D values were approximately constant (at 0.22–0.29) below 73 GPa. These values are somewhat lower than previous determinations (~ 0.4) in Al-bearing peridotite bulk compositions^{5–7} but are in good agreement with $K_D = 0.304 \pm 0.035$ at 25 GPa in Al-free peridotite material⁶. The high K_D in Al-bearing systems should be due to the high Fe³⁺ content of perovskite (see, for example, ref. 9). On the other hand, the value of K_D suddenly dropped to 0.07 ± 0.02 at 76 GPa (Fig. 2a). It then remained almost constant at 0.06–0.08 up to 159 GPa.

In order to explore the cause of strong Fe-enrichment in partial melt at pressures above 76 GPa, we collected the X-ray emission spectra of (Mg_{0.95}Fe_{0.05})SiO₃ glass at 300 K at increasing pressure, from 8 to 85 GPa (Fig. 3). At low pressures, the Fe K β' satellite peak was clearly observed at 7,045 eV, showing the presence of high-spin Fe²⁺ in the glass sample. The satellite peak diminished somewhat at 59 GPa, and vanished at 77 GPa. This indicates a spin collapse in ferrous iron. Insofar as the glass is a good analogue for the liquid state, such a high-spin to low-spin crossover of iron may also occur in melt at a similar pressure range, thus providing an explanation for the measured jump in Fe-enrichment in partial melt (see, for example, refs 10, 11). Indeed, the pressure range of the spin crossover observed in our glass matches the pressure where K_D changed dramatically (Fig. 2a). Our melts had higher Mg/Si ratios and FeO contents than (Mg_{0.95}Fe_{0.05})SiO₃ glass; however, the calculated mean Mg–O and Si–O coordination numbers for Mg₂SiO₄ liquid¹² are very similar to those for MgSiO₃ liquid¹³ at lower mantle pressures. Therefore, the spin crossover pressure, which depends on Mg/Fe–O coordination, would not shift significantly following a change in Mg/Si ratio in melt from 1 to 2. Theory^{14,15} suggests that the Fe content does not change the pressure range of the spin crossover when the Fe concentration is small (less than $\sim 20\%$), because Fe–Fe interactions are negligible. Multi-anvil experiments¹⁶ have shown a very sharp change in Fe partitioning between perovskite and ferropericlase at the commencement of the spin crossover, which is compatible with the present observations.

A strong change in iron partitioning suggests that partial melt becomes more dense at depths below 1,800 km. We calculated the

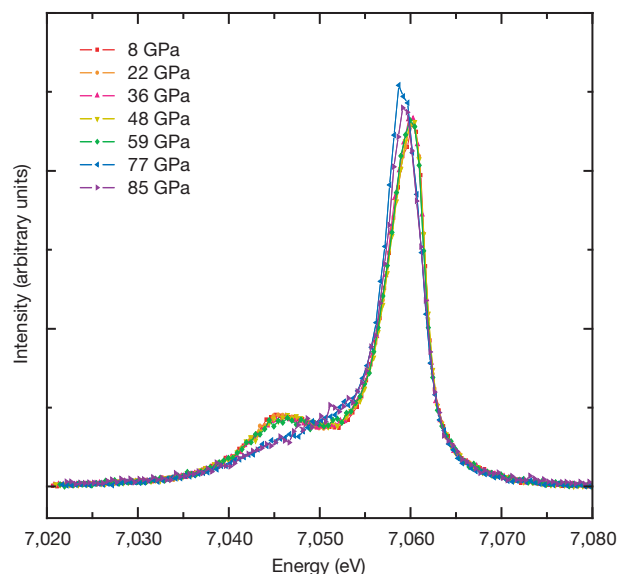


Figure 3 | Evolution of X-ray emission spectra of (Mg_{0.95}Fe_{0.05})SiO₃ glass with increasing pressure. Measurements were conducted at 300 K. All spectra are normalized to transmitted intensity, and shifted so that the weighted average of main (K β) plus satellite (K β') emission lines is set to 7,058 eV. The satellite peak decreased slightly at 59 GPa and completely disappeared at 77 GPa, indicating the spin crossover of iron.

density of (Mg,Fe)SiO₃ liquid in equilibrium with (Mg_{0.92}Fe_{0.08})SiO₃ perovskite—a typical composition for pyrolytic lower mantle^{16,17}—at 4,000 K as a function of pressure (Fig. 2b). For simplicity, we used $K_D = 0.25$ below 75 GPa and $K_D = 0.07$ at higher pressures. The molar volume of MgSiO₃ liquid was obtained by recent first principles calculations⁴, and the effect of iron is assumed to be the same for liquid and solid (perovskite)^{4,18}. Whereas (Mg,Fe)SiO₃ melt is buoyant compared to any of the typical lower-mantle minerals below 75 GPa, it suddenly becomes more dense at higher pressures. The difference from the preliminary reference Earth model (PREM)¹⁹ reaches 8% at the base of the mantle. Partial melts obtained in the present experiments had a higher Mg/Si ratio than MgSiO₃ (Supplementary Fig. 2). According to previous shock-wave compression experiments⁸, such high-(Mg+Fe)/Si melt is likely to be denser than the (Mg,Fe)SiO₃ melt considered above, suggesting that the density crossover between melt and solid may occur even at depths shallower than 1,800 km. Although details of the formation of melts at the base of the mantle are not yet clear, the present results provide a constraint on the plausible thickness of any stable melt layer of around 1,000 km (Fig. 4).

Labrosse *et al.*³ have presented a model in which dense melts produced soon after Earth's formation comprised a potentially large 'basal magma ocean' (BMO), which underwent slow crystallization over billions of years, at a rate governed by relatively sluggish solid state convection in the overlying solid mantle. The present experimental results provide a new physical rationale for the gravitational stability of a BMO up to ~1,000 km thick underneath the solid mantle, and the maximum plausible thickness inferred above is broadly consistent with the BMO hypothesis. For example, a BMO of ~1,000 km thickness would comprise about one-quarter of the mantle's mass; fractional crystallization and sequestering of incompatible heat producing elements (for example, U, Th) in the residual liquid as the layer cools would therefore account for the 'missing' chondritic complement of these species expected to be sequestered in a reservoir inside the mantle²⁰. The present results also allow us to make further predictions about the nature of the BMO and its chemical evolution through time.

As noted above, our results demonstrate that (Mg,Fe)SiO₃ perovskite is the first phase to crystallize from melts with a wide range of (Mg+Fe)/Si ratios; this is true even for Si-poor melt with (Mg+Fe)/Si ≈ 2.5 at the conditions of the core–mantle boundary (Supplementary Fig. 2). This strongly suggests that an episode of perovskite crystallization might well characterize the majority of the freezing history of a BMO, with other phases such as (Mg,Fe)O magnesiowüstite only crystallizing relatively late. In addition, perovskite crystals forming in the BMO would have been relatively depleted in iron and floated to the top of magmas at depths greater than 1,800 km (Fig. 4), because of the small K_D value for Fe/Mg partitioning. As a consequence, we can predict that the residual magma would evolve towards a FeO-rich/SiO₂-poor composition (that is, near the composition of wüstite), becoming more dense with time and probably also retaining a variety of incompatible volatile species.

Evolution through fractional crystallization as described above would also have affected the composition of cumulates that formed from the BMO. In particular, cumulates should become more Fe-rich with time, and presumably more dense, as they crystallize from an increasingly Fe-rich magma. These dense cumulates themselves might eventually become stable against complete entrainment by mantle convection currents, and pile up at the base of the mantle to form thermo-chemical piles (Fig. 4). The accumulation of dense solid material, which is around 2–3% denser than average mantle and which comprises ~2% of the total mantle volume, can explain the presence of two large, low-shear-wave velocity provinces (LLSVPs) at the base of the mantle beneath the Pacific and Africa²¹; and the expected Fe-enriched composition of the dense material is consistent with the magnitude of these anomalies²².

Larger degrees of crystallization in a BMO could also have left patches of dense residual mush in thin layers above the core–mantle

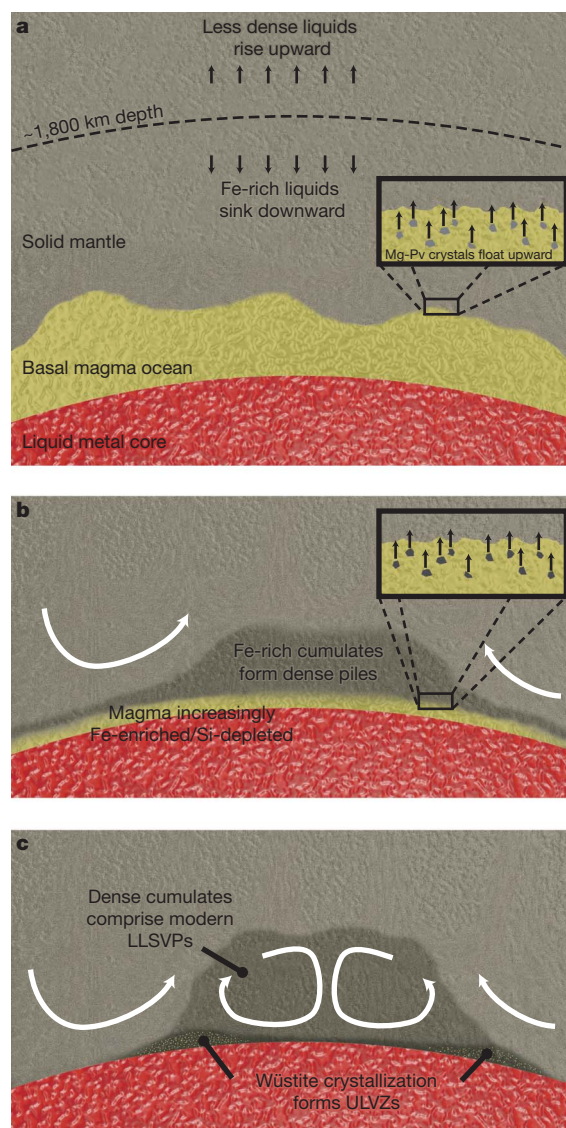


Figure 4 | Evolution and crystallization of dense melts in the deep mantle. **a**, During Earth's early history, any melts that form below ~1,800 km depth sink and accumulate at the base of the mantle, while any crystals that form owing to cooling of this dense magma will rise upward into the solid mantle. **b**, Fe-poor perovskite crystallization leaves a residual liquid enriched in FeO and depleted in SiO₂, and crystals forming from this evolved liquid may become dense enough to form thermo-chemical piles at the base of the solid mantle. **c**, The final stage of crystallization involves a composition close to wüstite, leaving behind a very dense thin layer that is consistent with the seismic properties inferred inside ULVZs. White arrows indicate schematic flow patterns in the convecting solid mantle.

boundary, which would explain the presence and seismic velocities of ultralow-velocity zones (ULVZs)²³. Such material would be maintained at or near the solidus over geological timescales because the residual liquid would sequester incompatible species that in turn depress the melting temperature. Alternatively, late crystallization of wüstite-rich cumulates from the Fe-rich/Si-poor melt could have left behind a dense solid layer with seismic properties that may also be consistent with seismic inferences for ULVZ, even in the absence of melt²⁴. The BMO model, together with the evolution of composition that we predict from our experiments, is compatible with both possibilities. Another possibility that might help to explain why ULVZs in

different regions do not always exhibit the same seismic signatures²⁵ is that inside some ULVZs the interstitial melt has drained out, leaving a largely melt-free solid rock strongly enriched in wüstite. In other ULVZs, owing to different dynamical conditions in the surrounding mantle, continual stirring of the mushy residue by viscous coupling to flows in the overlying mantle might have prevented compaction of the rock and removal of melt²⁶. The primary difference we would expect to observe between a melt-free, wüstite-rich ULVZ and a persistently mushy ULVZ would be the strength of the reduction in shear wave velocity, with a greater reduction inside mushy ULVZs²⁷.

METHODS SUMMARY

We loaded a platelet of $(\text{Mg}_{0.89}\text{Fe}_{0.11})_2\text{SiO}_4$ olivine starting material into the DAC together with argon pressure medium. After compression, the sample was heated by laser to induce melting. Subsequently it was recovered from the DAC and polished parallel to the compression axis. We obtained chemical compositions of the quenched liquid pool and the neighbouring solid phase (liquidus phase) from FE-EPMA analyses of these sections. In addition, several samples were further processed to a thin film and examined under a transmission electron microscope (TEM). The $\text{Fe}^{3+}/(\text{Fe}^{2+} + \text{Fe}^{3+})$ ratio of the quenched melt pocket was determined by electron energy-loss near-edge structure (ELNES) measurements. We also carried out the X-ray emission spectroscopy (XES) study of $(\text{Mg}_{0.95}\text{Fe}_{0.05})\text{SiO}_3$ glass in the DAC at high pressure and room temperature. The high-resolution spectra of the Fe K β line were collected through the cubic boron nitride + beryllium composite gasket, and indicate the change in the spin state of iron with increasing pressure.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 September 2010; accepted 16 February 2011.

Published online 24 April 2011.

- Stolper, E., Walker, D., Hager, B. H. & Hays, J. F. Melt segregation from partially molten source regions: the importance of melt density and source region size. *J. Geophys. Res.* **86**, 6261–6271 (1981).
- Agee, C. B. Crystal-liquid density inversions in terrestrial and lunar magmas. *Phys. Earth Planet. Inter.* **107**, 63–74 (1998).
- Labrosse, S., Hernlund, J. W. & Coltice, N. A crystallizing dense magma ocean at the base of the Earth's mantle. *Nature* **450**, 866–869 (2007).
- Stixrude, L. *et al.* Thermodynamics of silicate liquids in the deep Earth. *Earth Planet. Sci. Lett.* **278**, 226–232 (2009).
- Ito, E., Kubo, A., Katsura, T. & Walter, M. J. Melting experiments of mantle materials under lower mantle conditions with implications for magma ocean differentiation. *Phys. Earth Planet. Inter.* **143–144**, 397–406 (2004).
- Corgne, A. *et al.* Silicate perovskite-melt partitioning of trace elements and geochemical signature of a deep perovskitic reservoir. *Geochim. Cosmochim. Acta* **69**, 485–496 (2005).
- Lieske, C. *et al.* Compositional effects on element partitioning between Mg-silicate perovskite and silicate melts. *Contrib. Mineral. Petrol.* **149**, 113–128 (2005).
- Mosenfelder, J. L., Asimov, P. D. & Ahrens, T. J. Thermodynamic properties of Mg_2SiO_4 liquid at ultra-high pressures from shock measurements to 200 GPa on forsterite and wadsleyite. *J. Geophys. Res.* **112**, B06208, doi:10.1029/2006JB004364 (2007).
- McCammon, C. Perovskite as a possible sink for ferric iron in the lower mantle. *Nature* **387**, 694–696 (1997).
- Badro, J. *et al.* Iron partitioning in Earth's mantle: toward a deep lower mantle discontinuity. *Science* **300**, 789–791 (2003).
- Auzende, A.-L. *et al.* Element partitioning between magnesium silicate perovskite and ferropericlase: new insights into bulk lower-mantle geochemistry. *Earth Planet. Sci. Lett.* **269**, 164–174 (2008).
- de Koker, N. P., Stixrude, L. & Karki, B. B. Thermodynamics, structure, dynamics, and freezing of Mg_2SiO_4 liquid at high pressure. *Geochim. Cosmochim. Acta* **72**, 1427–1441 (2008).
- Karki, B. B. First-principles molecular dynamics simulations of silicate melts: structural and dynamical properties. *Rev. Mineral. Geochem.* **71**, 355–389 (2010).
- Tsuchiya, T., Wentzcovitch, R. M., da Silva, C. R. S. & de Gironcoli, S. Spin transition in magnesiowüstite in Earth's lower mantle. *Phys. Rev. Lett.* **96**, 198501 (2006).
- Bengtson, A., Persson, K. & Morgan, D. Ab initio study of the composition dependence of the pressure-induced spin crossover in perovskite $(\text{Mg}_{1-x}\text{Fe}_x)\text{SiO}_3$. *Earth Planet. Sci. Lett.* **265**, 535–545 (2008).
- Irifune, T. *et al.* Iron partitioning and density change of pyrolite in Earth's lower mantle. *Science* **327**, 193–195 (2010).
- Hirose, K. Phase transition in pyrolitic mantle around 670-km depth: implications for upwelling of plumes from the lower mantle. *J. Geophys. Res.* **107**, 2078, doi:10.1029/2001JB000597 (2002).
- Lundin, S. *et al.* Effect of Fe on the equation of state of mantle silicate perovskite over 1 Mbar. *Phys. Earth Planet. Inter.* **168**, 97–102 (2008).
- Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
- Hofmann, A. W. Mantle geochemistry: the message from oceanic volcanism. *Nature* **385**, 219–229 (1997).
- Hernlund, J. W. & Houser, C. On the distribution of seismic velocities in Earth's deep mantle. *Earth Planet. Sci. Lett.* **265**, 423–437 (2008).
- Cobden, L. *et al.* Thermochemical interpretation of 1-D seismic data for the lower mantle: the significance of nonadiabatic thermal gradients and compositional heterogeneity. *J. Geophys. Res.* **114**, B11309, doi:10.1029/2008JB006262 (2009).
- Williams, Q. & Garnero, E. J. Seismic evidence for partial melt at the base of Earth's mantle. *Science* **273**, 1528–1530 (1996).
- Wicks, J. K., Jackson, J. M. & Sturhahn, W. Very low sound velocities in iron-rich $(\text{Mg,Fe})\text{O}$: implications for the core-mantle boundary region. *Geophys. Res. Lett.* **37**, L15304, doi:10.1029/2010GL043689 (2010).
- Thorne, M. S. & Garnero, E. J. Inferences on ultralow-velocity zone structure from a global analysis of SPdKS waves. *J. Geophys. Res.* **109**, B08301, doi:10.1029/2004JB003010 (2004).
- Hernlund, J. W. & Jellinek, A. M. Dynamics and structure of a stirred partially molten ultralow velocity zone. *Earth Planet. Sci. Lett.* **296**, 1–8 (2010).
- Hier-Majumder, S. Influence of contiguity on seismic velocities of partially molten aggregates. *J. Geophys. Res.* **113**, B12205, doi:10.1029/2008JB005662 (2008).
- Fialin, M., Catillon, G. & Andrault, D. Disproportionation of Fe^{2+} in Al-free silicate perovskite in the laser heated diamond anvil cell as recorded by electron probe microanalysis of oxygen. *Phys. Chem. Miner.* **36**, 183–191 (2009).
- Komabayashi, T. *et al.* High-temperature compression of ferropericlase and the effect of temperature on iron spin transition. *Earth Planet. Sci. Lett.* **297**, 691–699 (2010).
- Ricolleau, A. *et al.* Density profile of pyrolite under the lower mantle conditions. *Geophys. Res. Lett.* **36**, L06302, doi:10.1029/2008GL036759 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Sinmyo for support with TEM analyses and K. Shimizu for preparing the glass sample. Discussion with R. Caracas and comments from D. Frost were helpful. C.-C. Chen is acknowledged for XES measurements at BL12XU Taiwan Beamline, SPring-8. Some of the melting experiments were conducted at BL10XU (SPring-8 proposal no. 2009B0087). J.H. was supported by the National Science Foundation (NSFEAR0855737).

Author Contributions R.N., H.O., S.T. and K.H. performed high-pressure experiments and chemical analyses. S.M. carried out the ELNES measurements. H.I. and N.H. supported the XES study at SPring-8. R.N., K.H. and J.H. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.H. (kei@geo.titech.ac.jp).

METHODS

Melting experiments at high pressure. High-pressure and high-temperature conditions were generated using laser-heated diamond-anvil cell (DAC) techniques. ($\text{Mg}_{0.89}\text{Fe}_{0.11}$) $_2\text{SiO}_4$ olivine from KLB-1 natural peridotite was used as a starting material³¹. The powdered sample was pressed into a disk with a typical thickness of 15–25 μm and was placed into a hole drilled at the centre of the rhenium gasket, which was pre-indented to a thickness of 25–40 μm . No additional metal absorber was used. Liquid argon was loaded cryogenically as a pressure medium. The sample was compressed with bevelled 120- μm and 150- μm , or flat 300- μm , culet diamond anvils, depending on a target pressure. Heating was performed with a fibre laser or a doughnut-mode Nd:YLF laser using a double-side heating technique that minimizes axial temperature gradients within the sample³². The laser spot size was about 20–30 μm on the sample.

The sample was heated for less than 1 s in order to avoid chemical segregation due to thermal diffusion under relatively large temperature gradient (see 'Attainment of chemical equilibrium' section in Supplementary Information). This heating duration was too short to allow accurate temperature measurement. Alternatively, we estimated the temperature in our experiment from the reported liquidus temperature of Mg_2SiO_4 (refs 8, 33) and the solidus temperature of natural peridotite³⁴, which give the upper and lower bounds, respectively (Supplementary Fig. 1). Pressure was determined from the Raman peak shift of the diamond anvil at 300 K after laser-heating³⁵. Here we assumed that the pressure was higher by 20% at melting temperature than at 300 K due to a contribution of thermal pressure, which was estimated from the previous *in-situ* X-ray diffraction measurements³⁶. The overall pressure uncertainty may be about $\pm 10\%$.

Chemical analyses of recovered samples. After complete pressure release, the sample was recovered from the DAC and glued on a silicon wafer using polymeric resin. Subsequently it was polished parallel to the compression axis by an Ar ion beam using an Ion Slicer (JEOL EM-09100 IS)³⁷. This method minimizes the damage of the sample surface, allowing us to observe microstructures within the extremely small sample. The section was examined by FE-EPMA (JEOL JXA-8500F) using an acceleration voltage of 10 kV and a beam current of 12 nA. We collected X-ray mappings for Si, Mg and Fe for each sample (Fig. 1 and Supplementary Figs 3 and 4). Chemical compositions of the quenched liquid pool and the neighbouring solid phase (ferropericlase, perovskite, or post-perovskite, depending on pressure) were obtained with defocused (5- μm size) and focused (<2- μm spatial resolution) beams, respectively (Supplementary Table 1). Synthetic wollastonite (CaSiO_3), periclase (MgO), and haematite (Fe_2O_3) were used as standards. In runs 21 and 24, however, the analyses of ferropericlase were contaminated by the melt pocket and were therefore estimated by subtracting the melt composition from the EPMA analyses until the SiO_2 content in ferropericlase becomes zero.

In addition, the samples in runs 17, 18 and 21 were further processed to a thin film by using an Ion Slicer and examined under a transmission electron microscope (TEM). For run 18, the $\text{Fe}^{3+}/(\text{Fe}^{2+} + \text{Fe}^{3+})$ ratio of the quenched melt pocket was determined by electron energy-loss near-edge structure (ELNES) spectroscopy measurements (Supplementary Fig. 5). The Fe $L_{2,3}$ -edge ELNES spectra were collected by using a JEM-2100M with an Enfi1000 spectrometer at Nagoya University. The measurements were made with a dispersion of 0.1 eV per channel, a typical energy resolution of about 0.8 eV, and an integration time of 25–100 s. The spectra were recorded using a macrosript for the Gatan Digital Micrograph (which controls the spectrometer) that corrects for energy drifts during data accumulation. This avoids the problem of peak broadening that follows prolonged data accumulation³⁸. The incident beam current was about 0.1 nA, and the fluence rate was about $50 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$. The quantitative determination of the $\text{Fe}^{3+}/$

($\text{Fe}^{2+} + \text{Fe}^{3+}$) ratio is based on the white line intensities at the Fe $L_{2,3}$ -edge, following the method described in ref. 39. The validity of the present analytical procedures was checked on a couple of silicate glass standards, whose $\text{Fe}^{3+}/(\text{Fe}^{2+} + \text{Fe}^{3+})$ ratios had been previously determined by Mössbauer spectroscopy measurements to be 0.38 and 0.83 (ref. 40). Our ELNES analyses determined the $\text{Fe}^{3+}/(\text{Fe}^{2+} + \text{Fe}^{3+})$ ratios for these glass samples to be 0.40 ± 0.02 and 0.79 ± 0.09 . A systematic change in $\text{Fe}^{3+}/(\text{Fe}^{2+} + \text{Fe}^{3+})$ ratios was not found with increasing electron beam irradiation time, up to 100 s.

X-ray emission spectroscopy measurements. High resolution X-ray emission spectra (XES) of the Fe K β line were obtained for ($\text{Mg}_{0.95}\text{Fe}_{0.05}$) SiO_3 glass at the BL12XU beamline of SPring-8 (ref. 41). The glass sample was prepared by quenching melt from 1,931 K in a $\text{CO}_2\text{-H}_2$ gas-mixing furnace. The oxygen fugacity in the high-temperature furnace was controlled to be slightly above the iron–wüstite buffer, in which all Fe was present as ferrous iron. The FE-EPMA analyses confirmed that the sample was glass without any quench-crystals and had a chemical composition of ($\text{Mg}_{0.95}\text{Fe}_{0.05}$) SiO_3 as a consequence of partial loss of iron into the Pt basket that held the sample in the furnace. X-ray diffraction measurements also showed the starting material to be amorphous.

XES measurements were made at room temperature in the DAC with increasing pressure from 8 to 85 GPa (Fig. 3). We used a cubic boron nitride (inner) + beryllium (outer) composite gasket. The glass powder was loaded into a hole in the gasket, which was pre-indented to about 80 μm in thickness. In order to excite the sample through diamond, monochromatic X-rays at 11 keV were chosen as the incident beam. The incident X-ray beam was focused on and collimated to 15- μm size at the sample. A 1-m Rowland circle type spectrometer, equipped with 1-m spherical bent Ge(620) analyser, was used to acquire the emission spectra. These spectra were collected through the gasket in order to avoid a loss of intensity due to absorption by the diamond. The spectra covered the energy range from 7,020 to 7,080 eV with a resolution of 0.5 eV.

31. Takahashi, E. Melting of a dry peridotite KLB-1 up to 14 GPa: implications on the origin of peridotitic upper mantle. *J. Geophys. Res.* **91**, 9367–9382 (1986).
32. Shen, G., Mao, H. K. & Hemley, R. J. in *Advanced Materials '96* 149–152 (Proc. 3rd NIRIM International Symposium on Advanced Materials, 1996).
33. Presnall, D. C., Weng, Y.-H., Milholland, C. S. & Walter, M. J. Liquidus phase relations in the system MgO-MgSiO_3 at pressures up to 25 GPa — constraints on crystallization of a molten Hadean mantle. *Phys. Earth Planet. Inter.* **107**, 83–95 (1998).
34. Fiquet, G. *et al.* Melting of peridotite to 140 gigapascals. *Science* **329**, 1516–1518 (2010).
35. Akahama, Y. & Kawamura, H. High pressure Raman spectroscopy of diamond anvils to 250 GPa: method for pressure determination in the multimegabar pressure range. *J. Appl. Phys.* **96**, 3748–3751 (2004).
36. Ozawa, H. *et al.* Experimental study of reaction between perovskite and molten iron to 146 GPa and implications for chemically distinct buoyant layer at the top of the core. *Phys. Chem. Miner.* **36**, 355–363 (2009).
37. Tateno, S., Sinmyo, R., Hirose, K. & Nishioka, H. The advanced ion-milling method for preparation of thin film using Ion Slicer: application to a sample recovered from diamond-anvil cell. *Rev. Sci. Instrum.* **80**, 013901 (2009).
38. Sasano, Y. & Muto, S. Energy-drift correction of electron energy-loss spectra from prolonged data accumulation of low SNR signals. *J. Electron Microsc.* **57**, 149–158 (2008).
39. van Aken, P. A., Liebscher, B. & Styrsky, V. J. Quantitative determination of iron oxidation states in minerals using Fe $L_{2,3}$ -edge electron energy-loss near-edge structure spectroscopy. *Phys. Chem. Miner.* **25**, 323–327 (1998).
40. Jayasuriya, K., O'Neill, H. S. C., Berry, A. J. & Campbell, S. J. A Mössbauer study of the oxidation state of Fe in silicate melts. *Am. Mineral.* **89**, 1597–1609 (2004).
41. Jarrige, I., Cai, Y. Q., Ishii, H., Hiraoka, N. & Bleuzen, A. Thermally activated charge transfer in a Prussian blue derivative probed by resonant inelastic x-ray scattering. *Appl. Phys. Lett.* **93**, 054101 (2008).

Evolution and metabolic significance of the urea cycle in photosynthetic diatoms

Andrew E. Allen^{1,2}, Christopher L. Dupont¹, Miroslav Oborník³, Aleš Horák³, Adriano Nunes-Nesi^{4†}, John P. McCrow¹, Hong Zheng¹, Daniel A. Johnson¹, Hanhua Hu^{2†}, Alisdair R. Fernie⁴ & Chris Bowler²

Diatoms dominate the biomass of phytoplankton in nutrient-rich conditions and form the basis of some of the world's most productive marine food webs^{1–4}. The diatom nuclear genome contains genes with bacterial and plastid origins as well as genes of the secondary endosymbiotic host (the exosymbiont⁵)^{1,6–10}, yet little is known about the relative contribution of each gene group to diatom metabolism. Here we show that the exosymbiont-derived ornithine-urea cycle, which is similar to that of metazoans but is absent in green algae and plants, facilitates rapid recovery from prolonged nitrogen limitation. RNA-interference-mediated knockdown of a mitochondrial carbamoyl phosphate synthase impairs the response of nitrogen-limited diatoms to nitrogen addition. Metabolomic analyses indicate that intermediates in the ornithine-urea cycle are particularly depleted and that both the tricarboxylic acid cycle and the glutamine synthetase/glutamate synthase cycles are linked directly with the ornithine-urea cycle. Several other depleted metabolites are generated from ornithine-urea cycle intermediates by the products of genes laterally acquired from bacteria. This metabolic coupling of bacterial- and exosymbiont-derived proteins seems to be fundamental to diatom physiology because the compounds affected include the major diatom osmolyte proline¹² and the precursors for long-chain polyamines required for silica precipitation during cell wall formation¹¹. So far, the ornithine-urea cycle is only known for its essential role in the removal of fixed nitrogen in metazoans. In diatoms, this cycle serves as a distribution and repackaging hub for inorganic carbon and nitrogen and contributes significantly to the metabolic response of diatoms to episodic nitrogen availability. The diatom ornithine-urea cycle therefore represents a key pathway for anaplerotic carbon fixation into nitrogenous compounds that are essential for diatom growth and for the contribution of diatoms to marine productivity.

An ornithine-urea cycle (OUC) driven by mitochondrial carbamoyl phosphate synthase (CPS) was long assumed to have originated in metazoans^{13–16} and is known to have facilitated key physiological and life-history adaptations in vertebrates^{13–15,17,18}. Genome sequence data for marine diatoms, however, suggested a much earlier origin for the OUC^{6,7,19}. Extended phylogenetic analysis and estimates of the timing of CPS diversification confirm that stramenopiles and haptophytes, which diverged much earlier than metazoans, also contain OUC-type CPS (novel CPS-a; Fig. 1). Additionally, it is evident that the CPS gene passed through two duplication events. The first duplication (1,181–1,500 million years ago; Fig. 1) pre-dates the emergence of metazoans in the eukaryotic tree (Fig. 1 and Supplementary Fig. 1). The second duplication (1,126–1,402 million years ago; Fig. 1) occurred before the diversification between metazoans, stramenopiles and haptophytes of OUC-type CPS (Fig. 1, clade A). Therefore, this duplication probably preceded the secondary endosymbiotic event(s) and associated plastid acquisitions in the latter two lineages. The

pattern and timing of CPS duplication and divergence indicate that the secondary endosymbiotic host cell of stramenopile and haptophyte algae shared a common ancestry with metazoans at the point of key metabolic innovations that are absent in the Plantae (Archaeplastida) lineage.

In addition to the expansion of this OUC CPS clade to include basal metazoans such as the sea anemone (*Nematostella vectensis*), two exclusively unicellular eukaryotic clades emerged. Novel CPS-a (Fig. 1) is related to mitochondrial OUC-type CPS enzymes which use ammonium (unCPS) and glutamine (ugCPS), and is found in Stramenopila and Haptophyta. Microscopy and measurements of protein activity in the pennate diatom *Phaeodactylum tricornutum* (Supplementary Figs 2 and 3 and Supplementary Table 1) support a mitochondrial localization for this gene product and indicate specificity for NH_4^+ rather than a preference for glutamine, as is the case in early diverging metazoans¹⁸. This indicates that the ancestral mitochondrial CPS was dependent on NH_4^+ rather than glutamine¹⁴.

The second new type of CPS, pgCPS2, thus named because it probably uses glutamine in pyrimidine synthesis, is found in stramenopiles, haptophytes and alveolates. This enzyme seems to be the result of the second CPS duplication (1,126–1,402 million years ago), which involved replacement of ancestral pgCPS1 with an enzyme related to mitochondrial unCPS. Although pgCPS2 is associated by phylogeny with OUC-related CPS, genes in the pgCPS2 clade do not encode a mitochondrial targeting signal and the proteins encoded are probably glutamine-dependent cytosolic enzymes that perform the first committed step of pyrimidine synthesis¹⁵. This is consistent with the finding that diatoms synthesize nucleotides in the cytosol and subsequently import them into plastids²⁰.

To evaluate the role of the diatom OUC, we investigated the timing and overall pattern of gene expression as well as the protein levels of unCPS and other OUC gene products in response to the addition of different nitrogen substrates to nitrogen-depleted *P. tricornutum* cultures. Genes involved in nitrate assimilation were used as a contrast. After nitrogen addition, these sets of genes showed unique but coordinated patterns of expression. In short (24 h) and long (120 h) nitrogen-recovery experiments (Fig. 2 and Supplementary Figs 4–6), genes involved in nitrate assimilation showed a strong upregulation that occurred specifically in response to nitrate and was followed by a quick decline, whereas OUC gene transcription occurred immediately after all nitrogen additions and remained elevated for longer than the nitrate-assimilation transcripts. Core OUC genes responded immediately (1–5 h), whereas downstream OUC-related genes responded on longer timescales (24–120 h), especially in response to NO_3^- . This may be due to the longer period of time required for nitrogen, and associated carbon, from NO_3^- assimilation to cycle through cellular pathways to the OUC. Furthermore, there were similar acute responses in the protein levels of mitochondrial unCPS to all nitrogen substrates

¹J. Craig Venter Institute, San Diego, California 92121, USA. ²CNRS UMR8197 INSERM U1024, Environmental and Evolutionary Genomics Section, Institute of Biology, Ecole Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France. ³Biology Centre ASCR, Institute of Parasitology and University of South Bohemia, Faculty of Science, Branišovská 31, 370 05 České Budějovice, Czech Republic. ⁴Max-Planck-Institut für Molekulare Pflanzenphysiologie, 14476 Potsdam-Golm, Germany. [†]Present addresses: Max Planck partner group Departamento de Biologia Vegetal, Universidade Federal de Viçosa, 36570-000 Viçosa, MG, Brasil (A.N.-N.); Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China (H.H.).

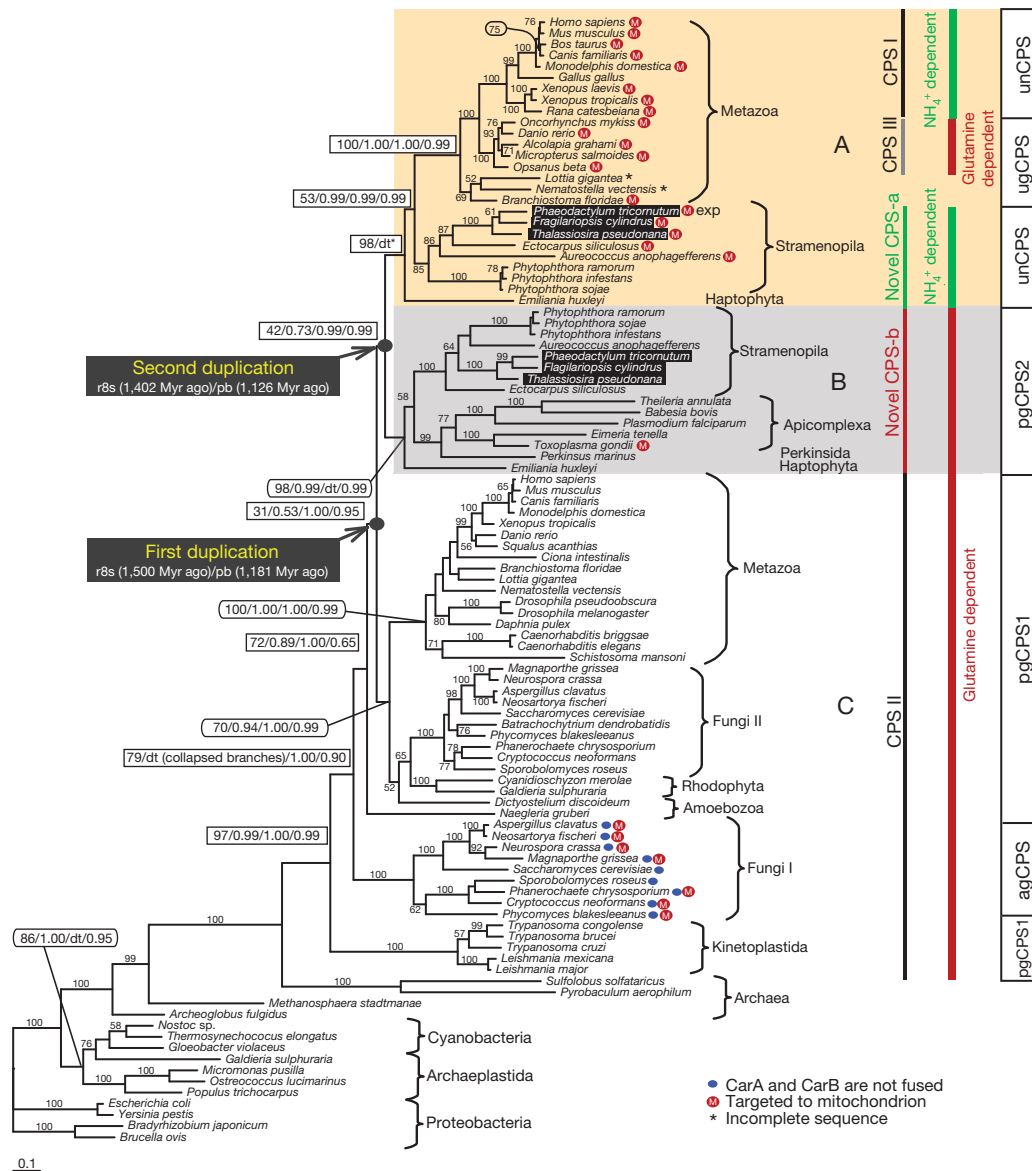


Figure 1 | Carbamoyl phosphate synthase phylogeny and divergence timing. Randomized accelerated maximum likelihood (RaXML with LG model) phylogenetic tree, inferred from amino acid sequences of carbamoyl phosphate synthases (CPS). Numbers above branches indicate RaXML bootstraps/PhyloBayes posterior probabilities/MrBayes posterior probabilities/PhyML. dt, different topology for particular method; exp, targeting was experimentally confirmed in this work. Sequences not composed of a fusion of CPS subunits CarA+B are marked. Abbreviations preceding 'CPS' indicate pathway (a, arginine synthesis; p, pyrimidine synthesis; u, urea cycle) and

and of nitrate reductase to nitrate (Supplementary Fig. 4). Urea protein levels, however, were consistently elevated with little variation in response to the source of nitrogen or the timing of the response to nitrogen addition (Supplementary Fig. 4), implying that there is near-constitutive intracellular production and turnover of urea regardless of the nitrogen substrate. Across 16 different conditions²¹, expression of OUC genes was elevated in response to conditions favourable for protein synthesis, such as the onset of light, elevated CO₂ and urea as the sole source of nitrogen (Supplementary Fig. 7). This implicates the diatom OUC in anabolic metabolism, in contrast to animal systems where the OUC typically has a catabolic role^{22,23}.

To gain further insight into the function of the diatom OUC, RNA interference (RNAi) techniques²⁴ were used to generate *P. tricornutum* cell lines with reduced levels of unCPS. unCPS protein content was reduced in six lines relative to wild type (Supplementary Fig. 8a). These

lines also showed a 15–30% reduction in steady-state growth rate when growing on NO₃⁻ or urea as the sole nitrogen source (Supplementary Fig. 8b). Additionally, when permitted to enter stationary phase, the unCPS RNAi lines required a longer period than wild-type cells to recover and resume growth upon transfer to fresh media, suggesting that impairment of mitochondrial unCPS significantly hampers rapid recovery from nitrogen limitation (Supplementary Fig. 8c). Within 48 h of exposure to fresh nitrogen, wild-type cells achieved maximum photosynthetic yield (variable fluorescence/maximum fluorescence, Fv/Fm), maximum growth rates and significant consumption of inorganic carbon. In contrast, the unCPS RNAi lines exhibited significant delays (Supplementary Fig. 9). These lags in exponential growth and photosynthetic yield resemble the physiological profile of algal cells treated with protein synthesis inhibitors during recovery from nitrogen limitation²⁵, indicating that mitochondrial CPS and the OUC

substrate (g, glutamine; n, ammonium). ugCPS and unCPS (clade A), pgCPS2 (clade B), and pgCPS1 (clade C) are indicated. CPS I, II and III designations have been used in the literature previously. The CPS nomenclature introduced here reflects proposed evolutionary chronology, substrate and pathway. The estimated times of both gene duplications leading to the appearance of the urea cycle are indicated: Myr, million years; Rates (r8s) 1.71 and PhyloBayes (pb) 3.2 were used to estimate divergence times. See Supplementary Fig. 1 and Supplementary Tables 3–5. Scale bar indicates 0.1 amino acid substitutions per position.

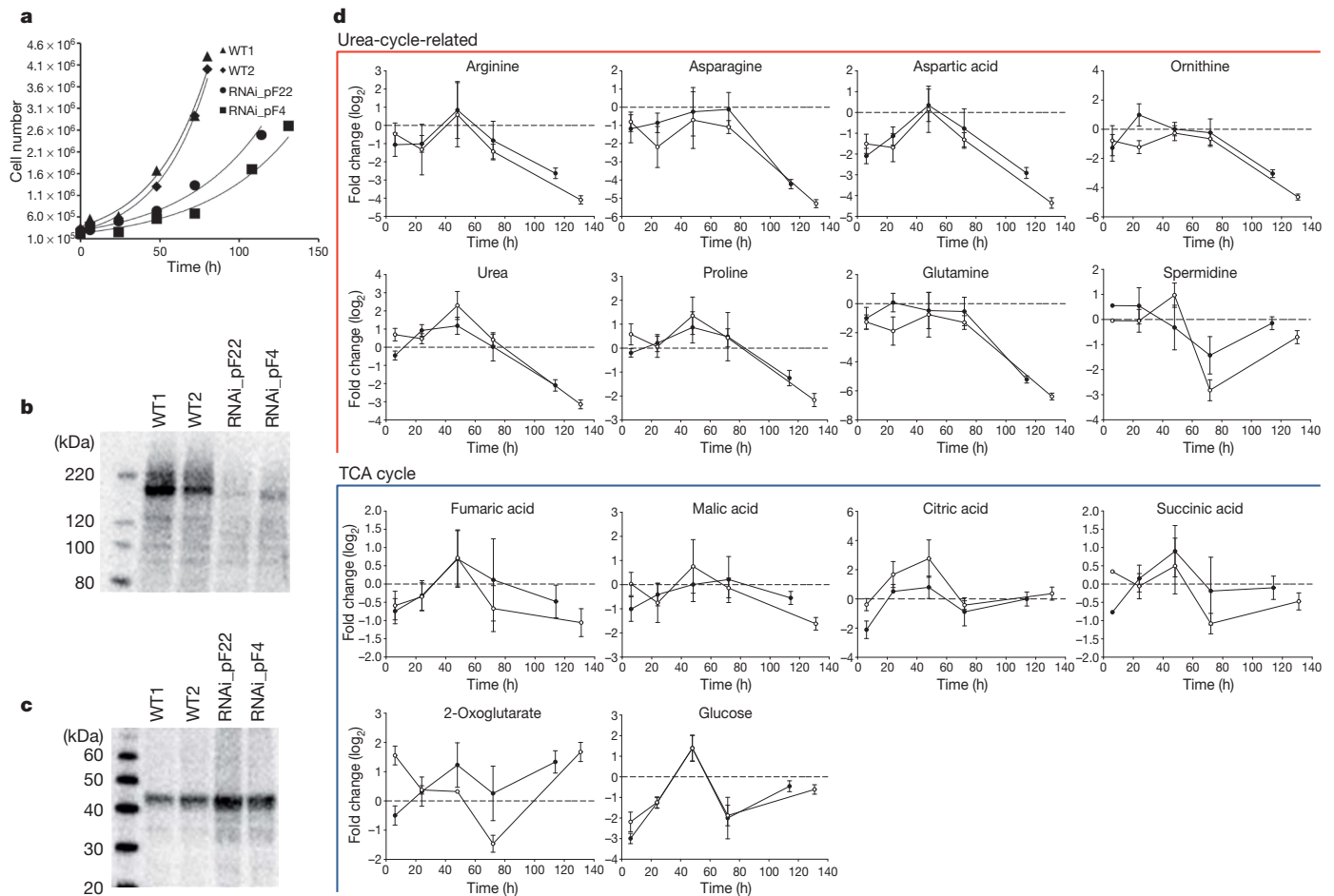


Figure 3 | Growth characteristics and metabolite abundance in wild-type and *unCPS* RNAi *P. tricornutum* lines. **a**, Growth curves of wild-type (WT1, WT2) and RNAi (pF22, pF4) lines. **b**, **c**, Western blot analyses of unCPS (**b**) and alternative oxidase (**c**) in wild-type and RNAi lines at $t = 24$ h. **d**, **e**, Fold changes in metabolites related to the urea cycle and citric acid cycle in two *unCPS* RNAi

lines relative to wild type. Wild-type metabolite levels were calculated as the mean of triplicate technical replicates performed on duplicate cultures ($n = 2$). Fold changes result from the comparison of metabolite levels at time points $t = 6$ h, 24 h, 48 h, 72 h and 80 h in wild-type lines with metabolites in the RNAi lines at time points $t = 6$ h, 24 h, 72 h, 114 h and 131 h. Error bars, s.d.

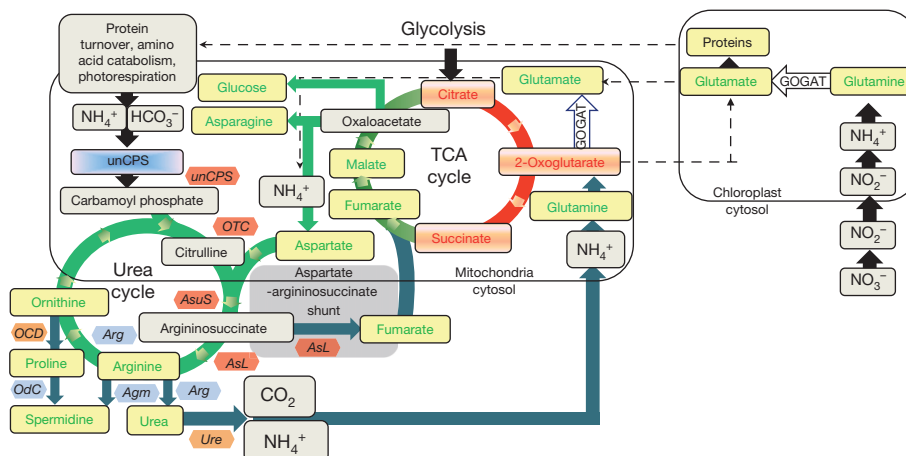


Figure 4 | Conceptual overview of the roles of unCPS and the diatom urea cycle on the basis of metabolite data from wild-type and RNAi lines. Metabolites depicted in green were significantly depleted (t -test, $P < 0.05$) whereas those shown in red were less affected in *unCPS* RNAi lines. Green and blue arrows indicate fluxes hypothesized to be particularly strongly affected by diatom *unCPS* impairment. Blue arrows further indicate potentially critical

control points that link the urea cycle to other major metabolic pathways. Core urea cycle genes that displayed coordinated expression, *AsL*, *AsuS*, *OTC* and *unCPS*, are indicated in red hexagons. *OCD* and *Ure*, depicted in orange hexagons, as well as *Arg*, *OdC* and to some extent *Agm*, depicted in blue hexagons, also showed independent coordination in overall expression. *GOGAT*, glutamate synthase.

key cellular metabolites may provide a partial explanation for the dominance of diatoms in the modern ocean.

METHODS SUMMARY

Cultures were grown in F/2 seawater with modified nitrogen substrates and levels. A full-length GST-CPS fusion product was purified from soluble *Escherichia coli* protein lysates using the MagneGST protein purification system (Promega). CPS activity was evaluated by monitoring the rate of conversion of [14 C]NaHCO₃ to [14 C]carbamoyl phosphate¹³. CPS RNAi lines were generated by construction of RNAi expression vectors with overlapping inverted repeat fragments (Supplementary Fig. 15) of the *P. tricornutum* CPS gene that were unique in the genome²⁴. Western blot analyses were performed with custom *P. tricornutum* genomic antibodies (Strategic Diagnostics) for CPS, alternative oxidase, nitrate reductase and urease. For quantitative reverse transcription PCR (qRT-PCR) experiments, 100–500 ng total RNA was reverse-transcribed and amplicons were quantified in qRT-PCR reactions. All PCR primers are given in Supplementary Table 2. Metabolites were extracted with methanol, derivatized and analysed by gas chromatography mass spectrometry²⁹. Maximum likelihood phylogeny was inferred from CPS amino acid sequences, computed using RAxML 7.2.6³⁰ under the LG + Gamma model of evolution. See Methods for details on estimates of CPS divergence times.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 December 2010; accepted 24 March 2011.

- Falkowski, P. G. *et al.* The evolution of modern eukaryotic phytoplankton. *Science* **305**, 354–360 (2004).
- Falkowski, P. G. & Oliver, M. J. Mix and match: how climate selects phytoplankton. *Nature Rev. Microbiol.* **5**, 813–819 (2007).
- Nelson, D. M., Treguer, P., Brzezinski, M. A., Leynaert, A. & Queguiner, B. Production and dissolution of biogenic silica in the ocean - revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Glob. Biogeochem. Cycles* **9**, 359–372 (1995).
- Smetacek, V. Diatoms and the ocean carbon cycle. *Protist* **150**, 25–32 (1999).
- Hamm, C. & Smetacek, V. in *Evolution of Primary Producers in the Sea* (eds Falkowski, P. G. & Knoll, A. H.) (Academic Press, 2007).
- Allen, A. E., Vardi, A. & Bowler, C. An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Curr. Opin. Plant Biol.* **9**, 264–273 (2006).
- Bowler, C. *et al.* The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–244 (2008).
- Moustafa, A. *et al.* Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* **324**, 1724–1726 (2009).
- Bowler, C., Vardi, A. & Allen, A. E. Oceanographic and biogeochemical insights from diatom genomes. *Ann. Rev. Mar. Sci.* **2**, 333–365 (2010).
- Allen, A. E. *et al.* Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl Acad. Sci. USA* **105**, 10438–10443 (2008).
- Kröger, N. & Poulsen, N. Diatoms—from cell wall biogenesis to nanotechnology. *Annu. Rev. Genet.* **42**, 83–107 (2008).
- Krell, A., Funck, D., Plettner, I., John, U. & Dieckmann, G. Regulation of proline metabolism under salt stress in the psychrophilic diatom *Fragilariopsis cylindrus* (Bacillariophyceae). *J. Phycol.* **43**, 753–762 (2007).
- Anderson, P. M. Glutamine- and N-acetylglutamate-dependent carbamoyl phosphate synthetase in elasmobranchs. *Science* **208**, 291–293 (1980).
- Hong, J., Salo, W. L., Lusty, C. J. & Anderson, P. M. Carbamoyl-phosphate synthetase-III, an evolutionary intermediate in the transition between glutamine-dependent and ammonia-dependent carbamoyl-phosphate synthetases. *J. Mol. Biol.* **243**, 131–140 (1994).
- Mommsen, T. P. & Walsh, P. J. Evolution of urea synthesis in vertebrates: the piscine connection. *Science* **243**, 72–75 (1989).
- Lawson, F. S., Charlebois, R. L. & Dillon, J. A. R. Phylogenetic analysis of carbamoylphosphate synthetase genes: Complex evolutionary history includes an internal duplication within a gene which can root the tree of life. *Mol. Biol. Evol.* **13**, 970–977 (1996).
- Guppy, M. The hibernating bear: why is it so hot, and why does it cycle urea through the gut. *Trends Biochem. Sci.* **11**, 274–276 (1986).
- Holden, H. M., Thoden, J. B. & Rauschel, F. M. Carbamoyl phosphate synthetase: an amazing biochemical odyssey from substrate to product. *Cell. Mol. Life Sci.* **56**, 507–522 (1999).
- Armbrust, E. V. *et al.* The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. *Science* **306**, 79–86 (2004).
- Ast, M. *et al.* Diatom plastids depend on nucleotide import from the cytosol. *Proc. Natl Acad. Sci. USA* **106**, 3621–3626 (2009).
- Maheswari, U. *et al.* Digital expression profiling of novel diatom transcripts provides insight into their biological functions. *Genome Biol.* **11**, R85 (2010).
- Esteban-Pretel, G. *et al.* Vitamin A deficiency increases protein catabolism and induces urea cycle enzymes in rats. *J. Nutr.* **140**, 792–798 (2010).
- Lee, B. *et al.* In vivo urea cycle flux distinguishes and correlates with phenotypic severity in disorders of the urea cycle. *Proc. Natl Acad. Sci. USA* **97**, 8021–8026 (2000).
- De Riso, V. *et al.* Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res.* **37**, e96 (2009).
- Young, E. B. & Beardall, J. Photosynthetic function in *Dunaliella tertiolecta* (Chlorophyta) during a nitrogen starvation recovery cycle. *J. Phycol.* **39**, 897–905 (2003).
- Morris, S. M. Regulation of enzymes of the urea cycle and arginine metabolism. *Annu. Rev. Nutr.* **22**, 87–105 (2002).
- Nunes-Nesi, A., Fernie, A. R. & Stitt, M. Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. *Mol. Plant* **3**, 973–96 (2010).
- Parker, M. S., Mock, T. & Armbrust, E. V. Genomic insights into marine microalgae. *Annu. Rev. Genet.* **42**, 619–645 (2008).
- Lisec, J., Schauer, N., Kopka, J., Willmitzer, L. & Fernie, A. R. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nature Protocols* **1**, 387–396 (2006).
- Stamatakis, A. RAxML-VI-HP: maximum likelihood-based phylogenetic analyses with thousands of taxa and models. *Bioinformatics* **22**, 2688–2690 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Meichenin and C. Lichtlé for assistance with electron microscopy, J. C. Thomas for CPS purification and activity experiments, A. Falcatore for advice on RNAi constructs and A. Main for screening and evaluation of RNAi lines. This study was supported by the National Science Foundation (NSF-OCE-0722374, NSF-OCE-0727997, NSF-MCB-1024913) and JCVI internal funding to A.E.A., the European Commission Diatomics project and Agence Nationale de la Recherche (France) (C.B.) and the Czech Science Foundation (206/08/1423) (M.O.).

Author Contributions A.E.A. and C.B. designed the study. A.E.A. performed CPS localization, confocal microscopy, protein purification, activity, overexpression and other laboratory experiments. A.E.A. and C.L.D. designed nitrogen-recovery experiments and the physiological characterization of the RNAi as well as wild-type experiments, which were performed by H.Z. and C.L.D.. H.Z. generated and screened RNAi lines. H.Z. and H.H. performed long-term and short-term nitrogen-recovery and related qPCR experiments. D.A.J. ran qPCR reactions and assisted with analyses of qPCR data. M.O., A.H. and A.E.A. generated and analysed phylogenetic and molecular clock data. A.N.N. and A.R.F. performed metabolite profiling of samples collected from RNAi and wild-type cultures. J.P.M., C.L.D. and A.E.A. analysed qPCR, metabolite and western blot data in detail. A.E.A. wrote the paper with assistance from C.L.D., A.R.F., C.B. and M.O. All the authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.E.A. (aallen@jcvl.org).

METHODS

Culturing. All cultures were grown in F/2 media with modified nitrogen substrates and levels. Short-term qRT-PCR experiments were performed on cultures grown under 12 h:12 h light:dark cycles and all other experiments were conducted in constant light. For short-term qRT-PCR experiments (Fig. 2), cells from an exponential culture grown in standard F/2 media ($882 \mu\text{M NO}_3^-$) were collected, washed twice with nitrogen-free media and then inoculated at $4 \times 10^5 \text{ cells ml}^{-1}$ in media containing $50 \mu\text{M NO}_3^-$. Cell number was monitored with cell counts and after 10 days at stationary phase, nitrogen-starved cells were collected and washed twice with nitrogen-free media, then duplicate cultures were inoculated into media containing different nitrogen sources ($50, 100$ or $882 \mu\text{M NO}_3^-$, $75 \mu\text{M NH}_4^+$ or $75 \mu\text{M urea}$). Through a diel cycle, cells were collected for RNA extractions at $t = 0$ (the beginning of the 12 h:12 h light period), then at 3 h, 5 h and 24 h.

For longer-term nitrogen recovery experiments, quantitative PCR (qPCR) and western blot experiments (Supplementary Fig. 4), pre-culturing was in media containing $100 \mu\text{M NH}_4^+$ as the sole nitrogen source. PO_4^{3-} was adjusted to $37.5 \mu\text{M}$ to force cells into nitrogen limitation owing to a low nitrogen:phosphorus ratio. Cultures were maintained in HEPES-buffered media (1 mM , pH 8.1), agitated with internal stir bars and bubbled with air. After NH_4^+ concentrations declined to below detection (500 nM), the cultures reached stationary phase. After 9 days in stationary phase, Fv/Fm levels declined to approximately 0.1. Duplicate cultures then received additions of either $500 \mu\text{M NO}_3^-$, $500 \mu\text{M NH}_4^+$ or $250 \mu\text{M urea}$. RNA and protein samples were collected at $t = 0 \text{ h}$, 6 h, 24 h, 28 h and 120 h.

For all other nitrogen-recovery experiments, culture flasks were maintained on stir plates, bubbled with air, and pre-culturing was in media with urea as the sole nitrogen source. As above, urea was added to a final concentration of $50 \mu\text{M}$, resulting in a media nitrogen:phosphorus ratio of 10:3.7. After a period of time in stationary phase that varied according to the experiment, cells were collected, rinsed with nitrogen-free media and inoculated into media containing $880 \mu\text{M NO}_3^-$ as the sole nitrogen source. For physiological experiments designed to measure the impact of CPS impairment on the capacity to recover from nitrogen limitation (Supplementary Figs 8 and 9), the period of stationary phase for the pre-cultures ranged from 7 to 13 days. For metabolite-profiling experiments (Fig. 3 and Supplementary Figs 10–14), the duration of the light period for stationary phase pre-cultures was reduced to 3 days so that no discernable lag phase was apparent between RNAi and wild-type cultures. To sample comparable points in the growth curve, RNAi and wild-type cells in these experiments were exposed to a reduced period of stationary phase pre-culturing and the duration of the lag phase before the onset of exponential growth, in terms of cell number and fluorescence, was negligible. According to Fv/Fm, cell number and fluorescence, all of the time points sampled for all cell lines in these experiments were during exponential growth.

Steady growth rates for wild-type and RNAi lines were obtained from growth in media with either nitrate or urea as the sole nitrogen source. Acclimated specific growth rate (μ) values were estimated from *in vivo* chlorophyll-*a* fluorescence (10 AU , Turner Instruments). Maximum photochemical yields of photosystem II were determined with a Water-PAM fluorometer (Walz). Cell counts were performed by microscopy using a Sedgwick-Rafter slide or a Beckman Quanta flow cytometer.

Cloning, RNAi specificity and transgenic expression. *Phaeodactylum tricornutum* cDNA encoding full-length CPS was PCR-amplified and cloned into a TOPO pENTR vector (Invitrogen). A clone containing an error-free sequence was selected for Gateway recombination (Invitrogen) with a diatom C-terminal YFP pDONR vector³¹. The resulting expression vector was transformed into *P. tricornutum* by particle bombardment³². Transformants were then screened by PCR and used for confocal imaging and electron microscopy immunolocalization using an antiserum against GFP. CPS pENTR vectors were also recombined with a pDONR vector designed for N-terminal GST tagging and expression in BL21A *Escherichia coli* cells. CPS RNAi lines were generated by construction of RNAi expression vectors with overlapping inverted repeat fragments (Supplementary Fig. 15) of the *P. tricornutum* CPS gene, followed by transformation in *P. tricornutum*²⁴. The 457-base-pair gene fragment used in the RNAi expression cassette is specific for *P. tricornutum* *unCPS* and is not homologous to any other region of the *P. tricornutum* genome. All cloning primer sets are listed in Supplementary Table 2.

CPS activity. A full-length GST-CPS fusion product was purified from soluble *E. coli* protein lysates using the MagneGST protein purification system (Promega). CPS activity was evaluated by monitoring the rate of conversion of [^{14}C]NaHCO₃ to [^{14}C]carbamoyl phosphate. [^{14}C]carbamoyl phosphate formed after 30 min at 20°C was measured after conversion to [^{14}C] urea by boiling and addition of NH_4Cl , followed by removal of [^{14}C]NaHCO₃ and other ^{14}C -labelled anions by elution from an anion exchange resin¹³.

RNAi screening, protein levels, gene expression and metabolites. Western blot analyses were performed with custom *P. tricornutum* Genomic Antibodies (Strategic Diagnostics) for CPS, nitrate reductase, urease and alternative oxidase. Screening of RNAi lines for protein levels was performed on lines which initially showed a reduction

in growth on urea relative to wild-type. Subsequently, these lines were evaluated for CPS and alternative oxidase protein levels. Six lines that showed CPS deficiency did not show depleted alternative oxidase levels. For qPCR experiments, 100–500 ng total RNA was reverse transcribed using the Quantitect Reverse Transcription Kit (Qiagen) and amplicons were quantified in qRT-PCR reactions with Fast SYBR Green Master Mix (Applied Biosystems). qRT-PCR primer sets were evaluated for optimal efficiency and the $2^{-\Delta\Delta\text{CT}}$ method was used to estimate fold change in gene expression, normalized to three endogenous control genes, 18S rDNA, TATA binding protein and histone H4³¹. ΔCT values were obtained by subtracting the mean values of experimental genes from a mean of the control genes for each sample. Using a Welch approximation for unequal group variances, a *P* value was estimated on the basis of the *t*-distribution that resulted from a between-subjects *t*-test evaluating the control RNA ($t = 0$) relative to a given experimental RNA. qRT-PCR primers are listed in Supplementary Table 2. Metabolites were extracted with methanol, derivatized and analysed by gas chromatography mass spectrometry²⁹. Metabolites were identified by comparison with database entries of authentic mass-spectral and retention-index (MSRI) libraries³³. Hierarchical clustering of gene expression and metabolite data are according to the log₂ of gene expression fold changes and metabolite correlation coefficients, which were plotted using the 'heatmap' function in R³⁴ without additional scaling. Hierarchical clustering of the expression fold changes and metabolite correlation as level vectors are shown as dendrograms and are the basis for the ordering of experimental conditions (columns) and genes (rows).

Phylogenetic analysis. CPS amino acid sequences (Supplementary Table 5) were aligned using MAFFT 6³⁵. The alignment was edited in Seaview 4³⁶. We tested all sequences for deviation from average amino acid composition using a chi-squared-based homogeneity test as implemented in Tree-Puzzle 5.2³⁷. Because some taxa that were important for interpretation did not pass the homogeneity test (see Supplementary Table 3 for a list), we decided to perform phylogenetic analyses using both empirical and mixture models. The maximum likelihood (ML) topology was computed using RAXML 7.2.6³⁰ under the LG + Gamma model of evolution. The tree with the highest likelihood score was chosen from 200 independent analyses starting with randomized parsimony trees. Non-parametric bootstrap support was estimated from 500 iterations. In addition, ML analysis was performed using an empirical profile mixture model (C40) as implemented in PhyML-CAT^{38–40}. The topology with the highest likelihood score from 15 independent runs starting with randomized trees using both NNI and SPR swapping algorithms was selected. Owing to the considerable computational burden, we performed an approximate likelihood-ratio test on the 'best' tree (that is, aLRT)⁴¹ instead of more traditional non-parametric bootstrap analysis. We also analysed our dataset using a Bayesian approach; again, with both empirical and mixture models. For the empirical approach, we used MrBayes 3.1^{42,43} with two chains run under the WAG + Gamma model for two million generations; priors were set to defaults and the first 500,000 generations were omitted from tree reconstruction as a burn-in. Four chains were run until they reached convergence in Phylobayes 3.2d⁴⁴ to infer the topology using the mixture model combined with LG empirical exchange rates (CAT-LG model).

Divergence time estimation. The fit of the clock-like behaviour to the 'best' tree was tested and rejected by likelihood-ratio test as implemented in Tree-Puzzle 5.2. We therefore used algorithms that enable relaxation of the molecular clock to infer the date of duplication events from the phylogenetic tree. The penalized likelihood (PL) and non-parametric-smoothing (NPRS) algorithms implemented in R8S 1.71^{45,46}, as well as a log-normal auto-correlated relaxed clock model⁴⁷ implemented in Phylobayes 3.2, were used to estimate divergence times. Because the PL and NPRS yielded virtually the same results, we show only those of PL. The fossil calibration points compiled from^{48,49} are listed in Supplementary Table 4.

1. Siaut, M. *et al.* Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* **406**, 23–35 (2007).
2. Falcione, A., Casotti, R., Leblanc, C., Abrescia, C. & Bowler, C. Transformation of nonselectable reporter genes in marine diatoms. *Mar. Biotechnol.* **1**, 239–251 (1999).
3. Schauer, N. *et al.* GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett.* **579**, 1332–1337 (2005).
4. R Development Core Team. R: A language and environment for statistical computing. ISBN 3-900051-07-0 (R Foundation for Statistical Computing, 2008).
5. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
6. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010).
7. Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: a maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504 (2002).
8. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
9. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109 (2004).

40. Quang, L. S., Gascuel, O. & Lartillot, N. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* **24**, 2317–2323 (2008).
41. Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539–552 (2006).
42. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**, 754–755 (2001).
43. Ronquist, F. & Huelsenbeck, J. P. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
44. Lartillot, N., Leppae, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009).
45. Sanderson, M. J. A nonparametric approach to estimating divergence times in the absence of rate consistency. *Mol. Biol. Evol.* **19**, 1218–1231 (1997).
46. Sanderson, M. J. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* **19**, 101–109 (2002).
47. Thorne, J. L., Kishino, H. & Painter, I. S. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**, 1647–1657 (1998).
48. Berney, C. & Pawlowski, J. A molecular timescale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc. R. Soc. A/B* **273**, 1867–1872 (2006).
49. Benton, M. J. & Donoghue, P. C. J. Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24**, 26–53 (2007).

Novel pathway for assimilation of dimethylsulphoniopropionate widespread in marine bacteria

Chris R. Reisch¹, Melissa J. Stoudemayer³, Vanessa A. Varaljay¹, I. Jonathan Amster³, Mary Ann Moran² & William B. Whitman¹

Dimethylsulphoniopropionate (DMSP) accounts for up to 10% of carbon fixed by marine phytoplankton in ocean surface waters^{1,2}, producing an estimated 11.7–103 Tmol S per year³, most of which is processed by marine bacteria through the demethylation/demethiolation pathway⁴. This pathway releases methanethiol (MeSH) instead of the climatically active gas dimethylsulphide (DMS) and enables marine microorganisms to assimilate the reduced sulphur^{5–7}. Despite recognition of this critical microbial transformation for over two decades, the biochemical pathway and enzymes responsible have remained unidentified. Here we show that three new enzymes related to fatty acid β -oxidation constitute the pathway that assimilates methylmercaptopropionate (MMPA), the first product of DMSP demethylation/demethiolation, and that two previously unknown coenzyme A (CoA) derivatives, 3-methylmercaptopropionyl-CoA (MMPA-CoA) and methylthioacryloyl-CoA (MTA-CoA), are formed as novel intermediates. A member of the marine roseobacters, *Ruegeria pomeroyi* DSS-3, requires the MMPA-CoA pathway for MMPA assimilation and MeSH production. This pathway and the ability to produce MeSH from MMPA are present in diverse bacteria, and the ubiquitous SAR11 clade bacterium *Pelagibacter ubique* possesses enzymes for at least the first two steps. Analysis of marine metagenomic data indicates that the pathway is widespread among bacterioplankton in the ocean surface waters, making it one of the most important known routes for acquisition of reduced carbon and sulphur by surface ocean heterotrophs.

The global importance of DMSP lies in its availability as a carbon and sulphur source for marine microorganisms and as a precursor of the gas dimethylsulphide (DMS)⁸, the oceanic emission of which leads to the formation of cloud condensation nuclei and promotion of solar radiation backscatter⁹. Two competing pathways exist for the bacterial catabolism of DMSP, one releasing DMS and the other releasing methanethiol (MeSH). Only recently has there been progress on identifying the specific biochemical pathways and genes responsible for these transformations. To date, four genes have been identified that encode proteins which catalyse the cleavage reaction that releases DMS^{10–13}, and one gene (*dmdA*) has been identified that encodes the initial demethylase in the pathway to MeSH¹⁴. Following demethylation, the intermediate MMPA is further catabolized to the highly reactive volatile sulphur gas MeSH¹⁵. This demethiolation pathway has long been proposed to be either an elimination or reductive cleavage, producing either acrylate or propionate, respectively^{16,17}. An alternative proposal suggested that MMPA is catabolized similarly to β -oxidation of fatty acids^{18–20}. In either case, MeSH has not been shown to be a major product of DMSP metabolism as its rapid turnover results in very low concentrations in both culture-based (Supplementary Table 1) and environmental experiments¹⁵. Thus, this report is the first direct evidence that MeSH is indeed a major degradation product of DMSP in at least some organisms.

To elucidate the pathway of DMSP demethiolation in the marine roseobacter *Ruegeria pomeroyi* DSS-3, CoA-containing intermediates

were examined. A *dmdA*[−] strain, which is incapable of DMSP demethylation, was supplied with DMSP, and the wild type was given MMPA. Upon CoA extraction and HPLC separation, an unknown CoA-containing intermediate was highly abundant in cells fed MMPA, compared to the mutant strain, which could only use the cleavage pathway (Supplementary Fig. 1). Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry indicated that the mass of this unknown product was 870.137 (Supplementary Fig. 2a), consistent with the theoretical monoisotopic mass of a CoA-MMPA thioester (MMPA-CoA) of 870.137 [M+H]⁺. MMPA-CoA was synthesized chemically and shown to have the same molecular weight and chromatographic retention time as the compound isolated from cell extracts. When [1,2,3-¹³C] DMSP was used as the carbon source for a mutant in this pathway (see below), the mass of MMPA-CoA purified from cell extracts increased as expected to 873.147 [M+H]⁺ (Supplementary Fig. 2b), indicating that this compound was in fact a product of DMSP metabolism.

To identify the enzyme catalysing the production of MMPA-CoA, the native enzyme was purified from *R. pomeroyi* DSS-3 cell extracts. One of four proteins remaining after purification, as judged by silver-stained SDS-polyacrylamide gel electrophoresis (SDS-PAGE), was identified as a medium chain fatty-acid CoA ligase by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) analysis (Supplementary Table 2). This protein, now designated as 3-methylmercaptopropionyl-CoA ligase (DmdB), was encoded by gene SPO2045. Confirmation that the gene encoded the correct protein was obtained by cloning SPO2045 into the pET101 expression vector and expressing it in *Escherichia coli*. The recombinant *E. coli* strain possessed MMPA-CoA ligase activity, whereas the host strain alone did not. The enzymatic reaction consumed ATP and produced AMP (Supplementary Fig. 3a) and is presumed to have produced pyrophosphate as well.

To elucidate the next step of the pathway, MMPA-CoA-consuming activity in *R. pomeroyi* DSS-3 was examined by incubating crude cell extracts with MMPA-CoA and various redox cofactors. Upon the addition of the artificial electron acceptors phenazine methosulphate or ferrocenium hexafluorophosphate, MMPA-CoA was consumed and an unknown intermediate was produced, the molecular mass of which was 868.121 [M+H]⁺ (Supplementary Fig. 2c), exactly two hydrogen atoms less than MMPA-CoA. This mass was consistent with the theoretical monoisotopic molecular mass of methylthioacryloyl-CoA (MTA-CoA; Fig. 1).

Upon the addition of MTA-CoA to *R. pomeroyi* DSS-3 crude cell extract, MeSH and free CoA were released. The enzyme catalysing the release of MeSH was purified to electrophoretic homogeneity from cell extracts of MMPA-grown *R. pomeroyi* DSS-3 (Supplementary Fig. 4) and identified as SPO3805, an enoyl-CoA hydratase and a member of the crotonase superfamily (cd06558) (Supplementary Table 3). Immediately upstream in the *R. pomeroyi* genome was a gene annotated as acyl-CoA dehydrogenase. We proposed that this gene, SPO3804, encoded the enzyme catalysing the production of MTA-CoA. SPO3804 and SPO3805 were each cloned into the pET101 expression vector and

¹Department of Microbiology, University of Georgia, Athens, Georgia 30602, USA. ²Department of Marine Sciences, University of Georgia, Athens, Georgia 30602, USA. ³Department of Chemistry, University of Georgia, Athens, Georgia 30602, USA.

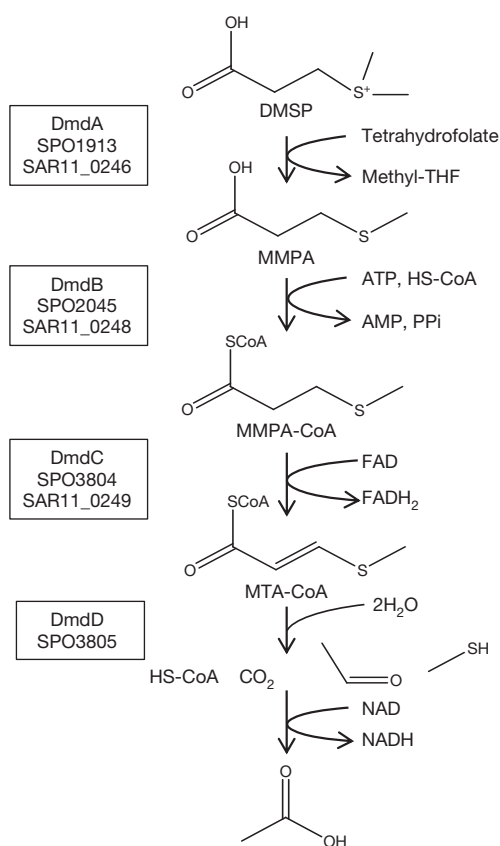


Figure 1 | Pathway of DMSP demethylation as identified in *R. pomeroyi* DSS-3. The genes identified thus far in both *R. pomeroyi* DSS-3 and *P. ubique* are indicated at each step. DMSP is first demethylated to MMPA in a tetrahydrofolate-dependent reaction catalysed by DmdA as described previously²⁵. A MMPA-CoA thioester is then formed in a reaction that consumes ATP and produces AMP. MMPA-CoA is then dehydrogenated forming an enoyl-CoA intermediate, MTA-CoA. In *R. pomeroyi* DSS-3, MTA-CoA is then hydrated by MTA-CoA hydratase in a reaction that releases the volatile sulphur product MeSH as well as free CoA, CO₂ and acetaldehyde. Acetaldehyde is then oxidized to acetic acid by acetaldehyde dehydrogenase.

expressed in *E. coli*. The recombinant *E. coli* strains had activity for MMPA-CoA dehydrogenase (SPO3804) and MTA-CoA hydratase (SPO3805), whereas the host strain alone did not (Supplementary Fig. 3b and 3c). Therefore, we designate these genes as *dmdC* and *dmdD*, respectively.

The DmdD-catalysed reaction produced stoichiometric amounts of free CoA and MeSH. Using ¹H and ¹³C NMR spectroscopy it was found that the three-carbon moiety was transformed into acetaldehyde (Supplementary Figs 5 and 6). The thioester bonded carbon of MTA-CoA was lost completely, presumably as CO₂, in the reaction (Supplementary Fig. 7). Quantification of acetaldehyde produced from MTA-CoA found 90% of the theoretical yield. *R. pomeroyi* DSS-3 is capable of further oxidizing acetaldehyde to acetate, based on the levels of acetaldehyde dehydrogenase activity (Table 1).

There are multiple lines of evidence that confirm the physiological significance of these activities. In extracts of chemostat-grown cells the levels of DmdB, DmdC and DmdD activities exceeded the minimum level, 57 nmol min⁻¹ per mg of protein, necessary to support growth (Table 1 and Methods). In addition, the amount of transcripts for *dmdB*, *dmdC* and *dmdD* increased during growth on MMPA or DMSP, as expected if the pathway was required for MMPA metabolism (Supplementary Fig. 8).

Mutations in each of the MMPA-CoA pathway genes also yielded phenotypes consistent with their participation in the pathway. A *dmdC*⁻ mutant (SPO3804::Tn5) could not grow on MMPA as the sole source of carbon (Supplementary Fig. 9), indicating that this pathway was essential for growth on MMPA. In contrast, the mutant grew similarly to wild-type with DMSP, indicating that the cleavage pathway, initiated by DddQ or DddP¹³, remained capable of supporting growth. Following growth with DMSP, the levels of DmdC and DmdD activity in the mutant were greatly reduced (Table 1). The low level of DmdC activity was consistent with the presence of additional *dmdC* orthologues in *R. pomeroyi* DSS-3 (see below). The low level of DmdD activity indicated that *dmdC* and *dmdD* were cotranscribed. This hypothesis was consistent with the absence of a recognizable promoter preceding *dmdD*, the coregulation of both genes, and the reduced levels of *dmdD* transcript in the *dmdC*⁻ mutant (Supplementary Fig. 8). Presumably, read-through of the transcriptional terminator on the kanamycin resistance marker was responsible for the low level of DmdD activity. The growth defect of the *dmdC*⁻ mutant was complemented with SPO3804 but not SPO3805 expressed on a plasmid, indicating that the failure of the *dmdC*⁻ mutant to grow on MMPA was not due to a polar affect on *dmdD*.

A *dmdD* (SPO3805::tet) mutant also failed to grow with MMPA, and growth on DMSP was severely inhibited (Supplementary Fig. 9d). Following growth on acetate, DmdD activity of the mutant was <0.5 nmol min⁻¹ mg⁻¹, whereas wild-type had activity of 12.7 nmol min⁻¹ mg⁻¹. Lastly, growth of a *dmdB* mutant (SPO2045::tet) was somewhat delayed during growth on MMPA (Supplementary Fig. 9e). Following growth with DMSP, the DmdB activity was reduced by only 40% compared to the wild-type (data not shown), which was consistent with the presence of an additional *dmdB* orthologue in *R. pomeroyi* DSS-3 (see below).

Homologues to the *R. pomeroyi* DSS-3 genes are abundant in the genomic database. The genomic database contained 36 bacteria that possessed *dmdA*, and all 36 also possessed *dmdB* and *dmdC*. However, the distribution of *dmdB* and *dmdC* was not limited to bacteria possessing *dmdA*, and many β and γ-proteobacteria as well as other bacteria not typically associated with marine systems possessed homologues with high sequence similarity to *dmdB* and *dmdC*. To confirm that *dmdA*-negative bacteria were capable of producing of MeSH from MMPA, pure cultures of representative bacteria were grown in the presence of MMPA. *Burkholderia thailandensis*, *Pseudomonas aeruginosa*, *Pseudoalteromonas atlantica*, *Myxococcus xanthus* and *Deinococcus radiodurans* all produced MeSH from MMPA. In contrast, *Escherichia coli*, which does not possess highly similar homologues, did not produce MeSH under the same conditions (Supplementary Table 4). Although a strain of *Burkholderia* was previously shown to possess a gene product capable of catalysing the DMSP cleavage reaction¹⁰, it is unlikely that many of the organisms possessing

Table 1 | MMPA-CoA pathway enzyme activities in cell extracts

Strain	Growth conditions	Enzyme assay*			
		MMPA-CoA ligase (DmdB)	MMPA-CoA dehydrogenase (DmdC)	MTA-CoA hydratase (DmdD)	Acetaldehyde dehydrogenase
<i>R. pomeroyi</i> DSS-3	2 mM MMPA (chemostat)	118 ± 17	1,226 ± 213	759 ± 36	132 ± 11
<i>R. pomeroyi</i> DSS-3	2 mM DMSP (batch)†	11 ± 1	281 ± 15	186 ± 18	34 ± 6
<i>R. pomeroyi</i> DmdC ⁻ (SPO_3804::Tn5)	2 mM DMSP (batch)†	27 ± 2	43 ± 5	12 ± 1	26 ± 3
<i>R. lacuscaerulensis</i>	1 mM MMPA (chemostat)	198 ± 30	39 ± 5	44 ± 13	n.d.

* Activities are reported as nmol min⁻¹ per mg of protein and are the result of triplicate experiments, ± s.d.

† The mutant and control wild-type strains were grown in batch culture to avoid selection for revertants in the chemostat.

n.d., not detected.

the MMPA-CoA pathway naturally encounter DMSP or MMPA derived from DMSP, and an alternative source of MMPA is likely. One possibility is that MMPA is derived from methionine, via the “off-pathway” reaction in the salvage pathway²¹. Regardless of the source of MMPA, the presence of the MMPA-CoA pathway in diverse bacteria further emphasizes its importance.

Both *dmdB* and *dmdC* are members of large gene families that encode enzymes with many different functions. To distinguish *dmdC* from closely related homologues with different functions, a selection of *dmdC* homologues from MeSH-producing bacteria were cloned and expressed in *E. coli*. Genes from *B. thailandensis*, *P. aeruginosa* and *Ruegeria lacuscaerulensis* encoded proteins with DmdC activity. Similarly, all three copies of *dmdC* from *R. pomeroyi* DSS-3 yielded functional proteins when expressed in *E. coli*. Lastly, SAR11_0249, the *dmdC* homologue from the SAR11 clade bacterium *Pelagibacter ubique* HTCC1062, was synthesized, expressed in *E. coli*, and shown to possess DmdC activity. These genes encompassed a well-defined clade within the acyl-CoA dehydrogenases (Fig. 2). Likewise, both copies of *dmdB* from *R. pomeroyi* DSS-3 and SAR11_0248, the *Pelagibacter ubique* HTCC1062 *dmdB* homologue, yielded functional

proteins when expressed in *E. coli*. These genes defined a similar DmdB clade of acyl-CoA ligases (Supplementary Fig. 10). Of the 49 sequenced genomes currently available in the JGI genomic database from the family *Rhodobacteraceae*, which contains the marine roseobacters, 47 and 49 possess *dmdB* and *dmdC* genes, respectively. The metabolism of reduced sulphur compounds in the marine roseobacters is complex, with some organisms producing MeSH from MMPA while lacking the ability to demethylate DMSP²². Thus, some roseobacters possess *dmdB* and *dmdC* even though they lack *dmdA* and are unable to demethylate DMSP. In addition to demonstrating that the MMPA-CoA pathway was widespread among marine bacteria, these experiments demonstrated that homologues are widespread in bacteria from a variety of habitats.

Given the abundance of *Pelagibacter* and the roseobacters in the ocean, both *dmdB* and *dmdC* are likely to be abundant in ocean surface waters. Analysis of the GOS metagenomic database²³ confirmed this hypothesis, as over 6200 homologues to DmdB and DmdC were found, indicating that these genes may be present in up to 61% of surface ocean bacterioplankton (Supplementary Table 5).

In contrast, the *P. ubique* homologue to the *R. pomeroi* DSS-3 DmdD possessed only low protein identity of 24%. When synthesized

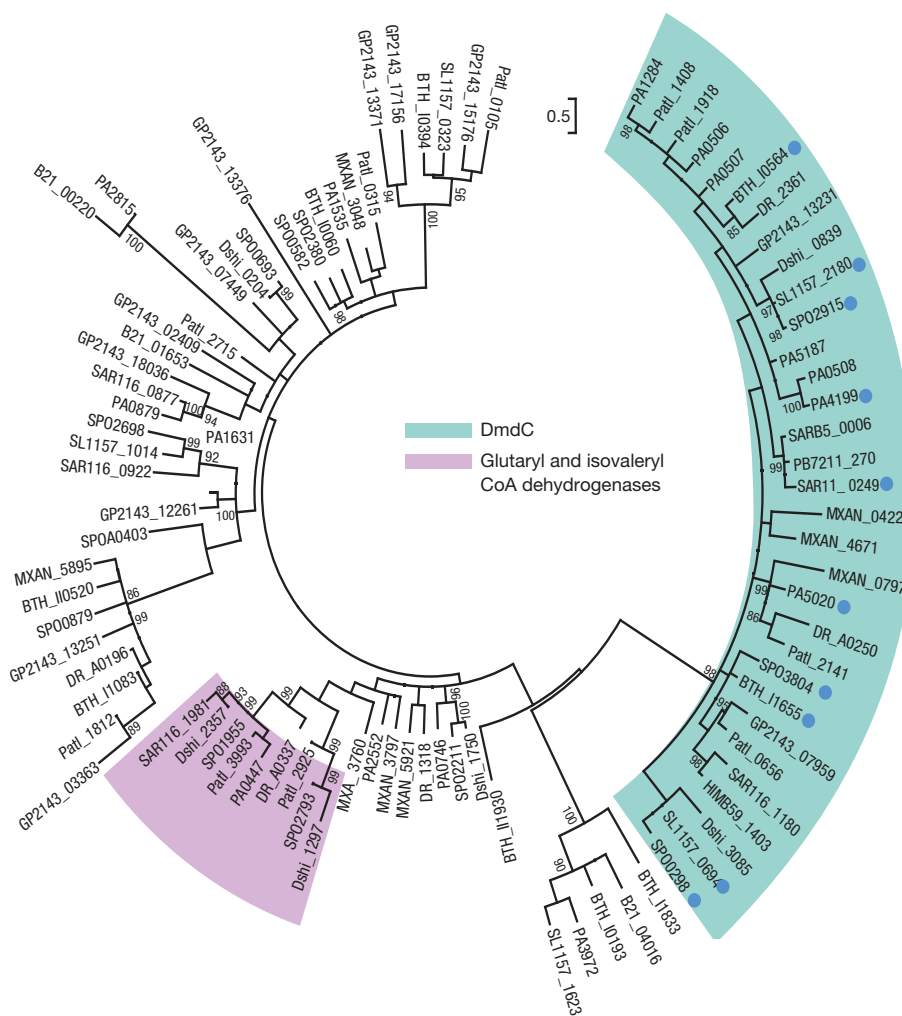


Figure 2 | Phylogenetic tree of DmdC from representative bacteria. Proteins from representative bacteria from marine surface waters as well as bacteria shown to produce MeSH from MMPA were included in the phylogenetic analysis. Proteins whose function was verified by recombinant expression in *E. coli* are indicated with blue dots. Methods for selecting homologous sequences are described in the supplementary methods. The original alignment contained 194 sequences, but for clarity, 101 out-group sequences were removed. Clusters with an identified or annotated function are labelled, whereas the remaining sequences possess conserved domains in the acyl-CoA dehydrogenase

superfamily (cl09966). Locus tags correspond to organisms as follows: *Escherichia coli* BL21(DE3) (B21), *Burkholderia thailandensis* (BTH), *Deinococcus radiodurans* (DR), *Dinoroseobacter shibae* (Dshi), marine γ -proteobacteria HTCC2143 (GP2143), SAR11 HIMB59 (HIMB59), *Myxococcus xanthus* (MXAN), *Pseudomonas aeruginosa* (PA), *Pseudoalteromonas atlantica* (Patl), *P. ubique* HTCC7211 (PB7211), *P. ubique* HTCC1062 (SAR11), *Puniceispirillum marinum* IMCC1322 (SAR116), SAR11 HIMB5 (SARB5), *R. lacuscaerulensis* (SL1157), *R. pomeroyi* (SPO).

and expressed in *E. coli*, it did not possess activity with MTA-CoA as the substrate. The gene SAR11_0247, annotated as an α - β fold hydrolase, was also cloned and expressed because it was located in between *dmdA* and *dmdB*. However, this gene product also did not have activity with MTA-CoA as the substrate. These results suggested that *dmdD* orthologues were not widely distributed and may have been replaced by non-orthologous isofunctional enzymes in some organisms. To investigate the pathway being used by *dmdD*-negative bacteria, the activity of the enzymes of the MMPA-CoA pathway were assayed in the *dmdD*-negative strain *Ruegeria lacuscaerulensis*. Cell extracts had activity for both DmdB and DmdC, as expected, but also for DmdD (Table 1). Therefore, a non-orthologous isofunctional enzyme may have replaced *dmdD* in this bacterium. Whether an isofunctional enzyme is also catalysing this step in *P. ubique* is unknown, but it is clear that orthologous proteins are not abundant in ocean surface waters as a BLASTp search against the GOS metagenomic database²³ yielded only 16 homologues with scores corresponding to an *e*-value of less than e^{-30} .

In conclusion, the novel MMPA-CoA pathway is widespread in marine and other bacteria. In oceans, this pathway leads to the formation of MeSH and acetate from the common osmolyte DMSP and prevents formation of the anti-greenhouse gas DMS. Acetate as a final product for the three-carbon moiety of DMSP is significant in that possible fates of acetate are numerous in cells. The DMSP cleavage pathway in a strain of *Halomonas* also resulted in the production of acetate as an end product, although the distribution and abundance of this pathway are unclear²⁴. The ecological function of the MMPA-CoA pathway outside the ocean is less understood. However, its ubiquity is strong evidence for an important role. Further investigations on the abundance, enzymology and expression of this pathway will be critical to our understanding of biosequestration and flux of reduced sulphur and carbon in marine and other ecosystems.

METHODS SUMMARY

R. pomeroyi DSS-3 and *R. lacuscaerulensis* were grown in an artificial seawater medium as described previously²⁵. Continuous cultures were grown in a carbon-limited chemostat with a dilution rate of 0.0416 h^{-1} and a doubling time of 24 h. To determine the minimum enzyme activity required to sustain growth in the chemostat, the dry weight of the chemostat culture was estimated using the attenuation at 660 nm (D_{660}) and the equation ($\mu\text{g ml}^{-1}$) = $364.74D_{660} + 6.7D_{660}$ (ref. 37). Using a substrate concentration of 1 mM for *R. lacuscaerulensis* or 2 mM for *R. pomeroyi* DSS-3 and a flow rate of 0.1 ml min^{-1} , there was a total of 100 or 200 nmol min^{-1} of substrate entering the chemostat. Assuming that 55% of dry weight was protein³⁸, the minimum enzyme activity required to sustain the observed MMPA consumption in the chemostat was estimated at 57 and 46 nmol $\text{min}^{-1}\text{ mg}^{-1}$ for *R. pomeroyi* DSS-3 and *R. lacuscaerulensis*, respectively. For full details of growth conditions, protein purifications, enzyme assays and genetic manipulations, see the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 October 2010; accepted 30 March 2011.

1. Archer, S. D., Widdicombe, C. E., Tarran, G. A., Rees, A. P. & Burkill, P. H. Production and turnover of particulate dimethylsulphoniopropionate during a coccolithophore bloom in the northern North Sea. *Aquat. Microb. Ecol.* **24**, 225–241 (2001).
2. Simó, R., Archer, S. D., Pedros-Alio, C., Gilpin, L. & Stelfox-Widdicombe, C. E. Coupled dynamics of dimethylsulphoniopropionate and dimethylsulfide cycling and the microbial food web in surface waters of the North Atlantic. *Limnol. Oceanogr.* **47**, 53–61 (2002).
3. Howard, E. C. *et al.* Bacterial taxa that limit sulfur flux from the ocean. *Science* **314**, 649–652 (2006).
4. Kiene, R. P., Linn, L. J. & Bruton, J. A. New and important roles for DMSP in marine microbial communities. *J. Sea Res.* **43**, 209–224 (2000).
5. van Duyl, F. C., Gieskes, W. W. C., Kop, A. J. & Lewis, W. E. Biological control of short-term variations in the concentration of DMSP and DMS during a *Phaeocystis* spring bloom. *J. Sea Res.* **40**, 221–231 (1998).

6. Ledyard, K. M. & Dacey, J. W. H. Microbial cycling of DMSP and DMS in coastal and oligotrophic seawater. *Limnol. Oceanogr.* **41**, 33–40 (1996).
7. Kiene, R. P., Linn, L. J., Gonzalez, J., Moran, M. A. & Bruton, J. A. Dimethylsulphoniopropionate and methanethiol are important precursors of methionine and protein-sulfur in marine bacterioplankton. *Appl. Environ. Microbiol.* **65**, 4549–4558 (1999).
8. Andreae, M. O. Ocean-atmosphere interactions in the global biogeochemical sulfur cycle. *Mar. Chem.* **30**, 1–29 (1990).
9. Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. Oceanic phytoplankton, atmospheric sulfur, cloud albedo and climate. *Nature* **326**, 655–661 (1987).
10. Todd, J. D. *et al.* Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. *Science* **315**, 666–669 (2007).
11. Curson, A. R. J., Rogers, R., Todd, J. D., Brearley, C. A. & Johnston, A. W. B. Molecular genetic analysis of a dimethylsulphoniopropionate lyase that liberates the climate-changing gas dimethylsulfide in several marine α -proteobacteria and *Rhodobacter sphaeroides*. *Environ. Microbiol.* **10**, 1099 (2008).
12. Todd, J. D., Curson, A. R., Dupont, C. L., Nicholson, P. & Johnston, A. W. The *dddP* gene, encoding a novel enzyme that converts dimethylsulphoniopropionate into dimethyl sulfide, is widespread in ocean metagenomes and marine bacteria and also occurs in some Ascomycete fungi. *Environ. Microbiol.* **11**, 1376–1385 (2009).
13. Todd, J. D. *et al.* DddQ, a novel, cupin-containing, dimethylsulphoniopropionate lyase in marine roseobacters and in uncultured marine bacteria. *Environ. Microbiol.* **13**, 427–438 (2010).
14. Howard, E. C., Sun, S. L., Biers, E. J. & Moran, M. A. Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environ. Microbiol.* **10**, 2397–2410 (2008).
15. Kiene, R. P. Production of methanethiol from dimethylsulphoniopropionate in marine surface waters. *Mar. Chem.* **54**, 69–83 (1996).
16. Taylor, B. F. & Gilchrist, D. C. New routes for aerobic biodegradation of dimethylsulphoniopropionate. *Appl. Environ. Microbiol.* **57**, 3581–3584 (1991).
17. Kiene, R. P. & Taylor, B. F. Demethylation of dimethylsulphoniopropionate and production of thiols in anoxic marine sediments. *Appl. Environ. Microbiol.* **54**, 2208–2212 (1988).
18. Taylor, B. F. & Visscher, P. T. in *Biological and Environmental Chemistry of DMSP and Related Sulfonium Compounds* (eds Kiene, R. P., Kirst, G. O., Keller, M. D. & Visscher, P. T.) Ch. 23, 265–276 (Springer-Verlag, 1996).
19. Bentley, R. & Chasteen, T. G. Environmental VOCs—formation and degradation of dimethyl sulfide, methanethiol and related materials. *Chemosphere* **55**, 291–317 (2004).
20. Tripp, H. J. *et al.* SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* **452**, 741–744 (2008).
21. Myers, R. W., Wray, J. W., Fish, S. & Abeles, R. H. Purification and characterization of an enzyme involved in oxidative carbon-carbon bond-cleavage reactions in the methionine salvage pathway of *Klebsiella pneumoniae*. *J. Biol. Chem.* **268**, 24785–24791 (1993).
22. Gonzalez, J. M., Kiene, R. P. & Moran, M. A. Transformation of sulfur compounds by an abundant lineage of marine bacteria in the α -subclass of the class Proteobacteria. *Appl. Environ. Microbiol.* **65**, 3810–3819 (1999).
23. Rusch, D. B. *et al.* The *Sorcerer II* Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol.* **5**, e77 (2007).
24. Todd, J. D. *et al.* Molecular dissection of bacterial acrylate catabolism—unexpected links with dimethylsulphoniopropionate catabolism and dimethyl sulfide production. *Environ. Microbiol.* **12**, 327–343 (2009).
25. Reisch, C. R., Moran, M. A. & Whitman, W. B. Dimethylsulphoniopropionate-dependent demethylase (DmdA) from *Pelagibacter ubique* and *Silicibacter pomeroyi*. *J. Bacteriol.* **190**, 8018–8024 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Wylie for assistance with NMR spectroscopy, S. Sharma, S. Sun and H. Luo for bioinformatics assistance, S. Gifford for technical advice, C. Smith and W. Crabb for technical assistance, and C. English for assistance with graphics. Funding for this research was provided by the National Science Foundation (MCB-0702125 and OCE-0724017) and the Gordon and Betty Moore Foundation.

Author Contributions C.R.R. performed growth experiments, enzyme assays, protein purifications, substrate synthesis, phylogenetic analysis, and all reaction analysis except MALDI-FT-ICR. M.J.S. and I.J.A. performed MALDI-FT-ICR analysis. C.R.R. and V.A.V. performed genetic modifications of *R. pomeroyi* DSS-3. V.A.V. performed RT-qPCR. C.R.R. and M.A.M. conducted bioinformatic analyses. C.R.R., M.A.M. and W.B.W. designed the experiments and wrote the paper. All authors reviewed the manuscript before submission.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to W.B.W. (whitman@uga.edu).

METHODS

Preparation of *R. pomeroyi* DSS-3 for intracellular acyl-CoA analysis. To investigate the biochemical pathways used to assimilate DMSP and MMPA in *R. pomeroyi* DSS-3, a targeted extraction of CoA-containing intermediates, such as those in the methylmalonyl-CoA or ethylmalonyl-CoA pathways, was performed. These experiments used a rich medium to grow the cells to a high density in order to obtain high amounts of CoA-containing intermediates which could be detected by ultraviolet absorbance at 260 nm after HPLC separation. Wild-type *R. pomeroyi* DSS-3, *dmdA*[−] (SPO1913::Tn5), or *dmdC*[−] (SPO3804::Tn5) was grown overnight in half-strength YTSS medium²⁶. Cells were collected, washed once, and resuspended in marine basal medium (MBM)²⁵ with 5 mM MMPA (wild-type) or DMSP (*dmdA*[−] and *dmdC*[−]). After overnight incubation, cells were again collected and resuspended in 3 ml of MBM with 5 mM DMSP or MMPA. Cells were incubated for 1 h, and then coenzyme-A containing intermediates were extracted.

Coenzyme-A extraction and analysis. Cells were quenched by the addition of trichloroacetic acid to 5%. Cell debris was removed by centrifugation at 10,000g for 10 min. The supernatant was then passed through an oligo purification cartridge (OPC, Applied Biosystems), which had been pre-washed with 70% acetonitrile/30% 10 mM KH₂PO₄ and then pre-equilibrated with 10 mM KH₂PO₄ (ref. 27). The OPC was then washed with 3 ml of 10 mM KH₂PO₄ and flushed with 1 ml of air. Products retained on the cartridge were eluted with 0.5 ml of 70% acetonitrile/30% 10 mM KH₂PO₄. The eluent was diluted to 1.5 ml with dH₂O, frozen, and lyophilized.

Products retained by the OPC were resolved by reverse phase chromatography using a 4.6 × 250 mm, 5 µm particle size, Aquasil column (Thermo Fisher). The column was developed at a flow rate of 1 ml min^{−1} with a gradient of 2–20% acetonitrile in 50 mM ammonium acetate (pH 6.0) over 40 min. Products were detected by absorbance at 260 nm. Identification of acyl-CoAs was based on the elution time of known standards. Additionally, some products observed on the ultraviolet trace did not contain coenzyme-A and were not further identified.

Preparation of *R. pomeroyi* DSS-3 cell extracts. A 9 l culture of *R. pomeroyi* DSS-3 was grown in a 15 l fermenter with MBM and 3 mM MMPA as the sole source of carbon. The air flow was set to 10 l min^{−1}. Cells were collected after 2 days of growth at an attenuation *D*₆₀₀ of 0.2 by centrifugation at 10,000g for 10 min and washed once with ice cold 50 mM Tris-HCl (pH 7.5). The pellet was resuspended in 5 ml of 50 mM Tris-HCl (pH 7.5). Cells were then lysed by passage through a French pressure cell at 100,000 kPa three times and centrifuged at 15,000g for 10 min. The supernatant was then centrifuged at 100,000g for 1 h at 4 °C. The supernatant at this stage was used for enzyme purifications.

In vivo production of MeSH. *Burkholderia thailandensis* E264, *Pseudoalteromonas atlantica* T6C and *Ruegeria lacuscaerulensis* were grown in MBM with 4 mM acetate with and without 1 mM MMPA. *Pseudomonas aeruginosa* PAO1 and *Escherichia coli* BL21(DE3) were grown in M9 minimal medium with 4 mM acetate or 4 mM acetate supplemented with 1 mM MMPA. *Deinococcus radiodurans*²⁸ and *Myxococcus xanthus*²⁹ were grown in defined medium as described previously, with and without 1 mM MMPA. All experiments were performed with 5 ml of medium in 28 ml Balch tubes sealed with a Teflon-coated stopper.

SPO3805 (DmdD) purification. Q-Sepharose HP chromatography: the cell extract was applied to a Q-Sepharose HP (GE Healthcare) anion exchange column (1.6 × 10 cm) equilibrated with 50 mM Tris-HCl (pH 8.0) at a flow rate of 2 ml min^{−1}. Protein was eluted with a gradient from 0–1 M NaCl over 8 column volumes. Activity eluted over 20 ml between 19 and 30 mS cm^{−1}.

Phenyl-Superose chromatography: active fractions from Q-Sepharose chromatography were pooled and made 1.7 M (NH₄)₂SO₄ by addition of solid (NH₄)₂SO₄. The solution was applied to a phenyl-Superose HR (GE Healthcare) hydrophobic interaction column (1 × 10 cm) at a flow rate of 1 ml min^{−1}, and the column was washed with 1 column volume of 1.7 M (NH₄)₂SO₄ in 50 mM Tris-HCl (pH 7.5). Protein was eluted with a gradient of 1.7–0 M (NH₄)₂SO₄ in 50 mM Tris-HCl (pH 7.5). Activity eluted at 110–96 mS cm^{−1}. Active fractions were pooled and concentrated with an Amicon Ultra centrifugal filter (10 kDa). The final concentrate was suspended in 50 mM potassium phosphate buffer (pH 7.5).

HiTrap Blue chromatography: the concentrated protein solution was then applied to a HiTrap Blue column (6 ml, GE Healthcare) that was equilibrated with 50 mM potassium phosphate (pH 7.5). The column was washed with 4 column volumes of buffer, and protein was eluted with a 0–2 M gradient of KCl in buffer over 6 column volumes. Activity eluted after start of the gradient. The 4 ml fraction containing the highest activity was concentrated using an Amicon Ultra centrifugal filter (10 kDa).

Sephacryl S200 chromatography: the protein concentrate was applied to a Sephacryl S200 (GE Healthcare) gel filtration column (1.6 × 25 cm) that was pre-equilibrated with 50 mM Tris-HCl (pH 7.5) and 150 mM NaCl. The protein was eluted with buffer at a flow rate of 1 ml min^{−1}. Fractions with activity were analysed on an SDS-PAGE gel and stained with GelCode Blue. The single protein

band was excised with a razor blade and analysed by in-gel trypsin digestion and MALDI-TOF mass spectrometry at the UGA PAMS facility.

Native SPO2045 (DmdB) purification. Q-Sepharose HP chromatography: the cell extract was applied and eluted from the column as described above. Activity eluted over 20 ml between 33 and 42 mS cm^{−1}.

Phenyl-Superose chromatography: active fractions from Q-Sepharose chromatography were pooled, made 1.7 M (NH₄)₂SO₄, and chromatographed as described above. Activity eluted between 39 and 22 mS cm^{−1}. Active fractions were pooled and concentrated with an Amicon Ultra centrifugal filter (10 kDa). Final concentrate was suspended in 2 ml of 50 mM potassium phosphate (pH 7.5).

HiTrap Blue chromatography: the concentrated protein solution was then applied to a HiTrap Blue column and eluted as described above. Activity eluted after the start of the gradient. The 4 ml fraction containing the highest activity was concentrated using an Amicon Ultra centrifugal filter (10 kDa).

Sephacryl S200 chromatography: the protein concentrate was applied to a Sephacryl S200 gel filtration column and chromatographed as described above. Active fractions were analysed in duplicate on two SDS-PAGE gels. One gel was stained with silver stain and the second with GelCode Blue. Protein bands from the GelCode-Blue-stained gel were excised with a razor blade and analysed by in-gel trypsin digestion and MALDI-TOF mass spectrometry at the UGA PAMS facility.

Preparation of recombinant *E. coli*. Genes were amplified from *R. pomeroyi* DSS-3, *R. lacuscaerulensis*, *Burkholderia thailandensis* E264 and *Pseudomonas aeruginosa* PAO1 genomic DNA or from pUC57 plasmids containing the synthesized *p. ubique* genes by PCR. PCR product was cloned into the pET101 expression vector using the methods recommended by Invitrogen, constructing the plasmids summarized in Supplementary Table 6. All constructs were made using TOP10 *E. coli* and then introduced into BL21(DE3) or Rosetta (DE3) cells for protein expression. Cells for protein expression were grown overnight in Luria-Bertani (LB) broth. The culture was used to inoculate a flask containing 250–1,000 ml of LB broth, which was incubated at 37 °C until reaching an attenuation of 0.5–0.6. Isopropyl-β-D-thiogalactoside (IPTG) was then added to a concentration of 0.2 mM, and the culture was incubated overnight at room temperature. The culture was then collected by centrifugation at 10,000g for 10 min, washed with 50 mM Tris-HCl (pH 7.5) and again centrifuged. The pellet was resuspended in 4 ml of buffer and lysed by passing twice through a French pressure cell at 100,000 kPa. Cell debris was removed by centrifugation at 15,000g for 10 min. The soluble fraction was subsequently used for protein purification or enzyme assays.

MALDI-FT-ICR. Mass spectra were collected on a 7 tesla BioApex Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometer (Bruker Daltonics) equipped with an intermediate pressure Scout100 MALDI source. Samples were overlain with a matrix consisting of saturated 2,5-dihydroxybenzoic acid.

Enzyme assays. HPLC analysis: analysis of DmdB, DmdC and DmdD reactions were performed using a 4.6 × 150 mm, 3 µm particle size, Hypersil Gold column (Thermo-Fisher) developed with a linear gradient of 2–20% acetonitrile in 50 mM ammonium acetate (pH 6) over 10 min.

Acetaldehyde dehydrogenase: acetaldehyde dehydrogenase activity was determined by measuring the increase in absorbance at 340 nm following the reduction of NAD to NADH using an extinction coefficient of 6,220 M^{−1} cm^{−1}. Enzyme assay mixture contained 50 mM HEPES (pH 7.5), 25 mM 2-mercaptoethanol, 1 mM NAD, 5 mM acetaldehyde, and 0.02–0.04 mg of cell extract. Reactions were initiated by the addition of acetaldehyde or cell extract.

Gas chromatography

Methanethiol was measured by headspace gas chromatography on an SRI 8610-C gas chromatograph with a Chromosil 330 column (Supelco) with N₂ carrier gas at a flow rate of 60 ml min^{−1}, an oven temperature of 60 °C, and a flame photometric detector. A standard curve for MeSH was obtained using a permeation tube (VICI Metronics).

Thiol quantification with DTNB. Free thiols resulting from the reaction of MTA-CoA were measured with 5,5'-dithio-bis(2-nitrobenzoic acid) (DTNB) using an extinction coefficient of 14,150 M^{−1} cm^{−1}. DTNB was prepared at a concentration of 10 mM in 50 mM potassium phosphate buffer (pH 7.5). A reaction mixture containing 50 mM potassium phosphate (pH 7.5), 10 µM MTA-CoA, 0.5 mM DTNB and 2 µg of purified DmdD were mixed, and the absorbance at 412 nm was taken.

Acetaldehyde quantification. Acetaldehyde was detected and quantified after reaction with semicarbazide to form acetaldehyde semicarbazone and HPLC separation. A reaction, containing 0.1 mM MTA-CoA, 10 mM semicarbazide and 50 mM potassium phosphate (pH 7.5) in 90 µl, was initiated with addition of 2 µg purified recombinant DmdD in 5 µl of buffer and run to completion. Five microlitres of 1 M formic acid were then added to quench the reaction and convert any acetaldehyde hydrate to acetaldehyde³⁰. The mixture was incubated at 37 °C for 2 h and centrifuged at 17,000g for 5 min. The supernatant was then analysed by

HPLC using a mobile phase of 2% acetonitrile with 0.5% acetic acid and ultraviolet detection at 224 nm (ref. 31).

Substrate synthesis. MMPA-CoA was synthesized by the mixed anhydride method³² or enzymatically using purified recombinant DmdB with the same conditions described above. MMPA-CoA that was synthesized by either method was purified by reverse phase chromatography using an Ultrasphere ODS preparative column (10 × 250 mm). The column was developed with 50 mM ammonium acetate (pH 6) and a gradient of 2–20% acetonitrile. MMPA-CoA was detected at 254 nm. Fractions containing MMPA-CoA were lyophilized, resuspended in dH₂O and again lyophilized.

MTA-CoA was synthesized enzymatically by the dehydrogenation of MMPA-CoA using purified recombinant SAR11_0249 with the same reaction conditions described above. MTA-CoA was then purified by reverse phase chromatography as described for MMPA-CoA.

¹³C-enriched substrates were synthesized enzymatically from ¹³C-enriched DMSP. ¹³C-enriched DMSP was synthesized as described previously³³ using ¹³C-enriched acrylic acid and dimethylsulphide. The ¹³C-enriched DMSP was then demethylated to MMPA using recombinant DmdA and tetrahydrofolate as the methyl acceptor as described previously²⁵. MTA-CoA was produced enzymatically from MMPA using purified DmdB and DmdC and purified as described above.

Nuclear magnetic resonance. ¹H NMR was performed on a Varian Unity Inova 500 MHz spectrometer. A total of 1,000 scans were recorded with a relaxation delay of 2 s and a pulse angle of 45°. ¹³C NMR was performed at 125 MHz with a 5 s relaxation delay, 45° pulse angle, and 2,000 scans.

Genetic modifications. Transposon mutagenesis was performed on *R. pomeroyi* DSS-3, mutants were screened for deficiency in MeSH production, and location of the transposon insertion was obtained by PCR and sequencing as described previously³.

Targeted gene replacements were made by introduction of *tetAR* into *dmdB* (SPO2045) and *dmdD* (SPO3805). Up and downstream regions of homology 1,000–1,500 base pairs in length and the *tetAR* genes from pRK415 were PCR-amplified and cloned into the pCR2.1 vector, which cannot replicate in *R. pomeroyi*, by sequence and ligation-independent cloning (SLIC)³⁴. Plasmid DNA was methylated by CpG methyltransferase as recommended by New England Biolabs and then introduced into *R. pomeroyi* DSS-3 cells by electroporation³⁵. Mutants were selected for ability to grow on tetracycline but not kanamycin, and confirmed by PCR.

Plasmids for complementation were made by PCR amplifying genes SPO3804, SPO3805, and both SPO3804 and SPO3805 from *R. pomeroyi* DSS-3 genomic DNA and ligating into the broad host expression vector pRK415 (ref. 36). The plasmids constructed are summarized in Supplementary Table 6.

Growth curves. Three-millilitre cultures of wild-type *R. pomeroyi* DSS-3 and mutant strains were grown in MBM with 2 mM glucose as the sole source of carbon. After overnight growth, the cultures were used to inoculate fresh 3 ml cultures, and attenuation was recorded at 600 nm.

Chemostat cultures. *R. pomeroyi* DSS-3 was grown at 30 °C in a carbon-limited chemostat using 2 mM MMPA in a total volume of 144 ml and a dilution rate of 0.0416 h⁻¹. *R. lacuscaerulensis* was grown at 33 °C with 1 mM MMPA with the same volume and dilution rates as above. A portion of the culture, 100 ml, was removed and immediately centrifuged at 10,000g for 10 min, washed with 1.5 ml of ice cold 50 mM HEPES (pH 7.5). The cell pellet was resuspended in 1 ml of 50 mM HEPES (pH 7.5). Cells were then lysed by bead beating for 2 min. Cell debris was removed by centrifugation at 17,000 for 5 min.

Volatile organic sulphur consumption. Chemostat cultures of *R. pomeroyi* DSS-3 grown with DMSP were collected and 2 ml was immediately placed into a 70 ml vial and sealed with a Teflon-coated stopper. Approximately 125 nmol of MeSH was then added to the vials from the MeSH permeation tube, and samples were equilibrated for 10 min at 30 °C. Headspace samples were then analysed by gas chromatography at 30 min, 1 h and 19 h.

Quantitative RT-PCR. RNA was extracted from chemostat-grown cell cultures of *R. pomeroyi* DSS-3 with acetate, MMPA, or DMSP as the sole carbon source. Cell culture, 50 ml, was collected into a 5 ml solution of 5% phenol in ethanol and centrifuged at 8,000g for 10 min. The supernatant was decanted, and pellets stored at -80 °C. RNA was extracted using the RNeasy mini kit (Qiagen), and DNA was removed with TURBO DNA-free kit (Ambion). The primer sequences, amplicon sizes, and annealing temperatures are listed in Supplementary Table 7.

Reactions were performed with the Bio-Rad One-Step RT-PCR reaction mix with SYBR Green, 300 nM primer concentrations, and 0.5 µl iScript reverse transcriptase (RT) enzyme. DNA plasmid standards were constructed from a PCR product for each gene using the pCR 2.1-TOPO vector (Invitrogen). RT-qPCR was carried out on a Bio-Rad iCycler iQ with the following cycling conditions: 50 °C for 10 min, 95 °C for 5 min, 40 cycles of 95 °C for 15 s and 61 °C for 1.0 min, and a final step of 95 °C for 1.0 min and 61 °C for 1.0 min, followed by a melt curve analysis. Transcript copies were calculated from standard curves and normalized per ng of RNA.

Phylogenetic analysis. The DmdB (SAR11_0249), DmdC (SAR11_0248) and RecA (SAR11_0641) from *P. ubique* were used as query sequences for a BLASTp search against the genome sequences of *Burkholderia thailandensis*, *Pseudomonas aeruginosa*, *Pseudomonas atlantica*, *Myxococcus xanthus*, *Deinococcus radiodurans*, *P. ubique* HTCC1062, HTCC1002, HTCC7211, HIMB5, HIMB59, marine γ-proteobacteria HTCC2143, *Puniceispirillum marinum* IMCC1322, *R. pomeroyi* DSS-3, *R. lacuscaerulensis* 1157, *Dinoroseobacter shibae* DFL-12, and *Escherichia coli* BL21(DE3). HIMB5 and HIMB59 were not included in the NCBI database at the time of analysis, the peptide sequences of the annotated genomes were obtained from JCVI. All hits with an expect value < e⁻¹⁰ were compiled into separate databases for DmdB and DmdC, and then the sequences were aligned using the MUSCLE algorithm in MEGA 5. Sequences with poor alignment and annotated as unrelated proteins were removed. Phylogenetic trees were built using maximum likelihood method in MEGA 5 with 100 bootstraps. Tree topology was confirmed with maximum likelihood method using PHYLIP v3.69.

GOS bioinformatics analysis. A custom reference database was established with DmdB, DmdC and RecA homologues from the bacteria listed above. Reference sequences were designated as orthologues if they clustered in a phylogenetic tree with authentic DmdBs, DmdCs or RecAs, or were otherwise considered paralogues. Separately, the DmdB (SAR11_0249), DmdC (SAR11_0248) and RecA (SAR11_0641) sequences from *P. ubique* were used as query sequences in BLASTp analysis against the GOS metagenomic database, and homologues with an expect value < e⁻²⁰ for DmdB and DmdC and < e⁻³ for RecA were retained. The GOS sequences were then blasted against the custom reference databases, and those with top hits to sequences in authentic clusters were summed while those with top hits to paralogous clusters were discarded. For DmdD, SPO3805 from *R. pomeroyi* DSS-3 was used as the query sequence for BLASTp analysis against the GOS metagenomic database.

26. Gonzalez, J. M. et al. *Silicibacter pomeroyi* sp. nov. and *Roseovarius nubinihibens* sp. nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments. *Int. J. Syst. Evol. Microbiol.* **53**, 1261–1269 (2003).
27. Deutsch, J., Grange, E., Rapoport, S. I. & Purdon, A. D. Isolation and quantitation of long-chain acyl-coenzyme-A esters in brain-tissue by solid-phase extraction. *Anal. Biochem.* **220**, 321–323 (1994).
28. Venkateswaran, A. et al. Physiologic determinants of radiation resistance in *Deinococcus radiodurans*. *Appl. Environ. Microbiol.* **66**, 2620–2626 (2000).
29. Bretscher, A. P. & Kaiser, D. Nutrition of *Myxococcus xanthus*, a fruiting myxobacterium. *J. Bacteriol.* **133**, 763–768 (1978).
30. Bell, R. P. & Higginson, W. C. E. The catalyzed dehydration of acetaldehyde hydrate, and the effect of structure on the velocity of protolytic reactions. *Proc. R. Soc. Lond. A* **197**, 141–159 (1949).
31. Gupta, N. K. A study of formaldehyde dismutation by liver alcohol dehydrogenase with NAD⁺-analogs. *Arch. Biochem. Biophys.* **141**, 632–640 (1970).
32. Stadtman, E. R. Preparation and assay of acyl coenzyme-A and other thiol esters - use of hydroxylamine. *Methods Enzymol.* **3**, 931–941 (1957).
33. Chambers, S. T., Kunin, C. M., Miller, D. & Hamada, A. Dimethylthetin can substitute for glycine betaine as an osmoprotectant molecule for *Escherichia coli*. *J. Bacteriol.* **169**, 4845–4847 (1987).
34. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nature Methods* **4**, 251–256 (2007).
35. Henriksen, J. R. *Physiology of dimethylsulfoniopropionate metabolism in a model marine Roseobacter, Silicibacter pomeroyi*. PhD thesis, Univ. Georgia (2008).
36. Keen, N. T., Tamaki, S., Kobayashi, D. & Trollinger, D. Improved broad-host-range plasmids for DNA cloning in Gram-negative bacteria. *Gene* **70**, 191–197 (1988).
37. Koch, A. L. in *Methods for general and molecular bacteriology* (eds Gerhardt, P., Murray, R. G. E., Wood, W. A. & Krieg, N. R.) 248–277 (American Society for Microbiology, 1994).
38. Neidhardt, F. C., Ingraham, J. L. & Schaechter, M. *Physiology of the bacterial cell: a molecular approach*. (Sinauer Associates, 1990).

A synthetic homing endonuclease-based gene drive system in the human malaria mosquito

Nikolai Windbichler¹, Miriam Menichelli¹, Philippos Aris Papathanos¹, Summer B. Thyme^{2,3}, Hui Li⁴, Umut Y. Ulge^{4,5}, Blake T. Hovde⁶, David Baker^{2,3,7}, Raymond J. Monnat Jr^{4,5,6}, Austin Burt^{1,8*} & Andrea Crisanti^{1,9*}

Genetic methods of manipulating or eradicating disease vector populations have long been discussed as an attractive alternative to existing control measures because of their potential advantages in terms of effectiveness and species specificity^{1–3}. The development of genetically engineered malaria-resistant mosquitoes has shown, as a proof of principle, the possibility of targeting the mosquito's ability to serve as a disease vector^{4–7}. The translation of these achievements into control measures requires an effective technology to spread a genetic modification from laboratory mosquitoes to field populations⁸. We have suggested previously that homing endonuclease genes (HEGs), a class of simple selfish genetic elements, could be exploited for this purpose⁹. Here we demonstrate that a synthetic genetic element, consisting of mosquito regulatory regions¹⁰ and the homing endonuclease gene *I-SceI*^{11–13}, can substantially increase its transmission to the progeny in transgenic mosquitoes of the human malaria vector *Anopheles gambiae*. We show that the *I-SceI* element is able to invade receptive mosquito cage populations rapidly, validating mathematical models for the transmission dynamics of HEGs. Molecular analyses confirm that expression of *I-SceI* in the male germline induces high rates of site-specific chromosomal cleavage and gene conversion, which results in the gain of the *I-SceI* gene, and underlies the observed genetic drive. These findings demonstrate a new mechanism by which genetic control measures can be implemented. Our results also show in principle how sequence-specific genetic drive elements like HEGs could be used to take the step from the genetic engineering of individuals to the genetic engineering of populations.

HEGs encode highly specific endonucleases with recognition sequences that typically occur only once per host genome, and have been identified in unicellular organisms in all three biological domains¹⁴. HEG-induced DNA double strand breaks (DSB) activate the recombinational repair system of the cell, which uses the homologous chromosome carrying the HEG as a template for repair. As a result the HEG is copied to the broken chromosome in a process referred to as 'homing'. HEGs use this transmission distortion mechanism to spread through populations¹⁵. To investigate *I-SceI* activity *in vivo* we have developed an experimental system consisting of three distinct transgenic mosquito lines, the Donor, the Reporter and the Target, carrying either the *I-SceI* gene or its recognition site at identical positions on homologous chromosomes (Supplementary Fig. 1). For this purpose we used an *A. gambiae* docking line¹⁶ that allowed the site-specific integration of three different plasmids carrying the red fluorescent protein (RFP) transformation marker on chromosome 3R (Supplementary Fig. 2). The Donor line was generated using the construct pHome-D, containing a 3×P3-GFP (green fluorescent protein) transcription unit interrupted by a synthetic HEG element

consisting of the *I-SceI* gene and the regulatory regions of the male testis-specific *A. gambiae* $\beta 2$ -tubulin gene¹⁰. The Reporter line was developed using the construct pHome-R, containing an *I-SceI* cleavage site that shifts out of frame the coding sequence of the *GFP* gene. The Reporter locus allows the scoring of *I-SceI* cleavage activity by monitoring the frequency of GFP⁺ individuals in which the GFP reading frame was restored via non-homologous end joining (NHEJ) in the progeny of Donor/Reporter trans-heterozygous males (Fig. 1a, b). Finally, the Target line was developed using pHome-T, containing the *I-SceI* cleavage site within the coding sequence of a functional *GFP* gene. This construct contains a diagnostic NotI recognition site that facilitates the molecular genotyping of homing events. The Target locus allows the assessment of *I-SceI* homing activity in the progeny of Donor/Target trans-heterozygous males crossed with wild-type females by measuring the increase in the frequency over a 1:1 ratio of GFP[−] to GFP⁺ individuals arising from the insertion of the HEG gene into the GFP open reading frame (Fig. 1d, e).

When Donor/Reporter trans-heterozygous females were crossed to wild-type males all progeny showed the expected GFP[−] phenotype, as the $\beta 2$ -tubulin promoter regulating *I-SceI* is not active in females. By contrast 3% of the progeny from Donor/Reporter trans-heterozygous males and wild-type females showed a GFP⁺ phenotype (Fig. 1c). Sequencing of PCR products from the region around the *I-SceI* site showed that in 5 out of 20 GFP⁺ individuals the correct reading frame had been restored by NHEJ repair events. The remaining 15 GFP⁺ mosquitoes showed in place of the *I-SceI* site a sequence that resembled the region joining the 3×P3 promoter and the *CFP* gene (cyan fluorescent protein), which lacks a unique restriction site present in the 3×P3-GFP cassette (Supplementary Fig. 2). We established from one such GFP⁺ individual the HEG-resistant Control strain, containing all three fluorescent marker genes but lacking the *I-SceI* site within the GFP sequence (Supplementary Fig. 1). The remaining 97% of the progeny from Donor/Reporter trans-heterozygous males and wild-type females were GFP[−] and the majority of these mosquitoes (93%) showed a GFP[−]RFP⁺CFP⁺ phenotype expected to arise either from an intact GFP[−] parental locus, NHEJ events that did not restore GFP expression or *I-SceI* homing events (Fig. 1c).

To test for the occurrence of homing we analysed the progeny of crosses between Donor/Target trans-heterozygous and wild-type mosquitoes (Fig. 1f). As expected, the ratio of GFP⁺:GFP[−] phenotypes in the offspring of Donor/Target trans-heterozygous females crossed to wild-type males was about 50:50. By contrast, in the reciprocal cross of trans-heterozygous males and wild-type females the ratio was 14:86. The excess of GFP[−] progeny, the majority of which were RFP⁺CFP⁺, could originate either from NHEJ events or as a result of homologous repair involving the HEG⁺ chromosome (that is, homing). To investigate the molecular nature of GFP inactivation we performed a PCR

¹Imperial College London, Department of Life Sciences, South Kensington Campus, London, SW7 2AZ, UK. ²Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA.

³Graduate Program in Biomolecular Structure and Design, University of Washington, Seattle, Washington 98195, USA. ⁴Department of Pathology, University of Washington, Seattle, Washington 98195, USA. ⁵Graduate Program in Molecular and Cellular Biology, University of Washington, Seattle, Washington 98195, USA. ⁶Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. ⁷Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. ⁸Imperial College London, Department of Life Sciences, Silwood Park Campus, Ascot, SL5 7PY, UK. ⁹Department of Experimental Medicine, University of Perugia, Via Del Giochetto, 06122 Perugia, Italy.

*These authors contributed equally to this work.

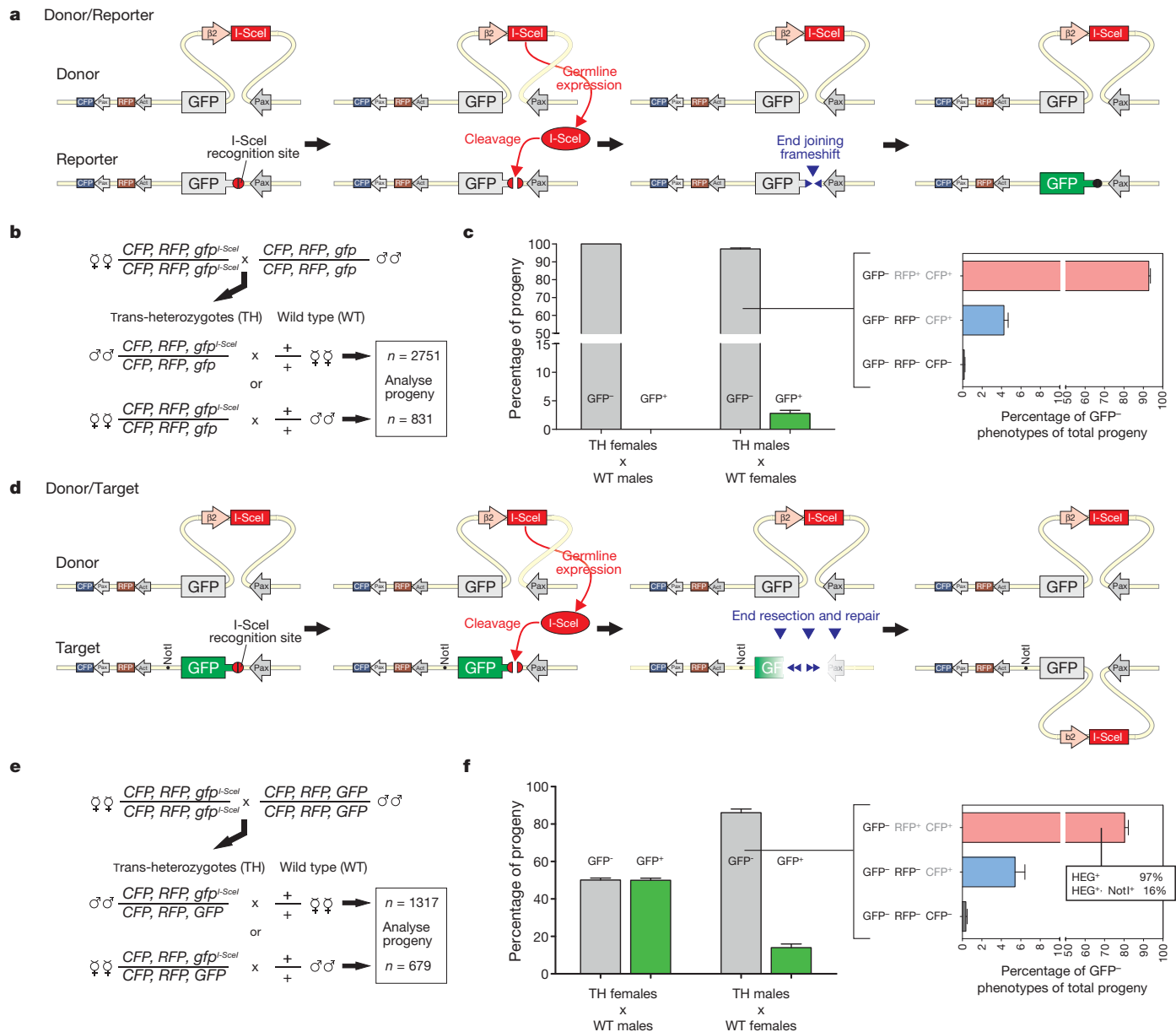


Figure 1 | Analysis of HEG activity in transgenic mosquitoes.

a, d, Anticipated molecular events unfolding in Donor/Reporter (**a**) or Donor/Target (**d**) trans-heterozygous (TH) males. The Donor locus expresses I-SceI under the control of the male germline promoter $\beta 2$ -tubulin. The Reporter and the Target loci contain an I-SceI recognition site within the *GFP* gene. **a**, In Donor/Reporter TH males I-SceI activity is detected by scoring events that restore the *GFP* reading frame upon cleavage of the I-SceI recognition site. **d**, In Donor/Target TH males cleavage of the Target locus is followed by end resection and homing of the I-SceI gene from the homologous chromosome. This leads to the inactivation of the *GFP* reporter gene and can also lead to co-

analysis of the region spanning the *GFP* locus and encompassing the I-SceI gene or its recognition site. The results showed that 97% of $GFP^- RFP^+ CFP^+$ individuals contained the HEG cassette (Fig. 1f). The estimated cleavage rate for I-SceI was therefore about 95%, and the overall homing rate 56%. Importantly, the diagnostic NotI marker present only on the Target locus allowed the identification of recombinant $GFP^- HEG^+ NotI^+$ chromosomes that were generated as a result of homing events (Supplementary Fig. 3). We were able to detect the NotI site, located ~0.7 kilobases from the I-SceI cleavage site, in 16% of HEG^+ chromosomes analysed, indicating that this marker was retained in 45% of all homing events (and lost due to co-conversion in the remaining 55%). In both sets of male trans-heterozygous to

conversion of the NotI molecular marker (Pax, 3×P3 promoter; Act, *Actin5C* promoter). **c, f**, Phenotypic analysis of progeny from crosses of Donor/Reporter (**b, c**) or Donor/Target (**e, f**) trans-heterozygote with wild-type (WT) mosquitoes. The column graphs show the percentage of GFP^- and GFP^+ individuals. The bar graphs on the right show, as a percentage of the total progeny, the $GFP^- RFP^+ CFP^+$, $GFP^- RFP^- CFP^+$ and $GFP^- RFP^- CFP^-$ individuals observed. The inset (**f**, right panel) shows the molecular genotype of $GFP^- RFP^+ CFP^+$ individuals analysed for the presence of the HEG and the NotI molecular markers.

wild-type crosses about 4–5% of the progeny were $GFP^- RFP^- CFP^+$, and a small number of larvae lacked all three visible markers (Fig. 1c, f). These phenotypes were not observed in progeny of trans-heterozygous females, suggesting that they were the result of I-SceI activity accompanied by deletions encompassing parts of the *RFP* gene or the entire locus. These experiments are summarized in Supplementary Table 1.

Another independent transgenic line, referred to as Ectopic Target, was generated by transposase-mediated integration of the pHome-T plasmid on chromosome 2 (Supplementary Fig. 1). When Donor/Ectopic Target trans-heterozygous males were crossed to wild-type females the frequency of the GFP^- phenotype in the progeny was 88%, compared to approximately 50% in the female trans-heterozygous

control cross (Supplementary Fig. 4). However none of 94 $\text{GFP}^- \text{RFP}^+ \text{CFP}^-$ individuals, the phenotypic class expected to contain non-parental HEGs, carried the HEG sequence. This experiment indicates that, in the absence of a repair template on the homologous chromosome, I-SceI cleavage activity does not induce detectable homing. Finally, we observed no significant deviation from a 1:1 ratio of GFP^- and GFP^+ progeny from crosses of trans-heterozygous mosquitoes in which the Donor locus was combined with the Control locus (data not shown).

To test whether the observed transmission ratio distortion allows for efficient genetic drive of I-SceI in receptive *A. gambiae* populations, we monitored its transmission dynamics in five cage populations of 600 individuals over 8 to 12 generations. Cage populations containing the I-SceI Target allele at initial frequencies of 90% or 50% were seeded with the I-SceI Donor allele at a frequency of 10% or 50%, respectively. GFP dominance results in an initial frequency of GFP^- individuals of 1% or 25% in the two experimental conditions. In subsequent generations GFP^+ individuals are expected to carry at least one allele of the original GFP^+ target gene or a misrepaired GFP^+ allele, whereas GFP^- individuals contain two alleles in which GFP has been inactivated either by insertion of the HEG or NHEJ. At each generation a random sample of the progeny was visually analysed for the GFP marker at the larval stage. In all populations the frequency of GFP^- individuals increased rapidly over time (Fig. 2). The frequency rose from about 1% to 60% in 12 generations (cage 1), and from about 1% to 40% in 10 generations (cage 2). In the two populations seeded with higher initial HEG frequencies GFP^- individuals reached about 75–80% after 8 generations. By contrast the frequency of GFP^- individuals did not change significantly in a population (cage 6) in which the HEG Donor line was used in combination with the non-receptive Control line (Fig. 2b), indicating that the absence of GFP expression in $\text{GFP}^- \text{RFP}^+ \text{CFP}^+$ mosquitoes did not result in a measurable fitness advantage over $\text{GFP}^+ \text{RFP}^+ \text{CFP}^+$ mosquitoes. We generated deterministic and stochastic population genetic models, using as parameters the experimentally derived rates of cleavage, homologous repair and NHEJ, assuming no fitness differences among genotypes (Supplementary Fig. 5a). The observed dynamics in the population cages fall well within the stochastic variation expected from the model (Fig. 2), indicating a quantitative match between the experimental data and our theoretical understanding of HEG transmission dynamics. If I-SceI had any effect on mosquito fitness, it was not large enough to significantly affect this concordance.

Detailed phenotypic and molecular analyses were carried out at different generations on individuals sampled from the mosquito population of cage 1. More than 90% of all GFP^- mosquitoes were $\text{RFP}^+ \text{CFP}^+$ for 12 generations (Supplementary Fig. 5b). To confirm that the rise of the $\text{GFP}^- \text{RFP}^+ \text{CFP}^+$ phenotype reflected a parallel increase in the HEG allele we performed a PCR assay on randomly chosen mosquitoes to determine the presence of the I-SceI gene. The frequency of individuals positive for the HEG cassette rose from about 19% to 86% by generation 12 (Fig. 2c). Moreover, NotI digests of the PCR products showed that the frequency of individuals with chromosomes carrying both the HEG and the NotI marker, a combination that was absent at the beginning of the experiment, increased to 50% by generation 9 (Fig. 2c). The dynamics of both HEG^+ and NotI^+ allele frequencies matched expectations from stochastic simulations (Fig. 2c). We conclude that the rise in the frequency of GFP^- individuals reflected the corresponding increase in the frequency of the HEG allele. The increase in the frequency of the NotI marker in the Donor allele pool indicates that homing is the cause for the observed rise in the frequency of HEG^+ individuals.

Our results demonstrate that homing can occur at appreciable frequencies in the germline of *A. gambiae* and therefore address a fundamental uncertainty that previously had been associated with proposals to use HEGs for pest control, namely whether HEGs would function in animals as they do in microbes. HEGs do not occur naturally in the

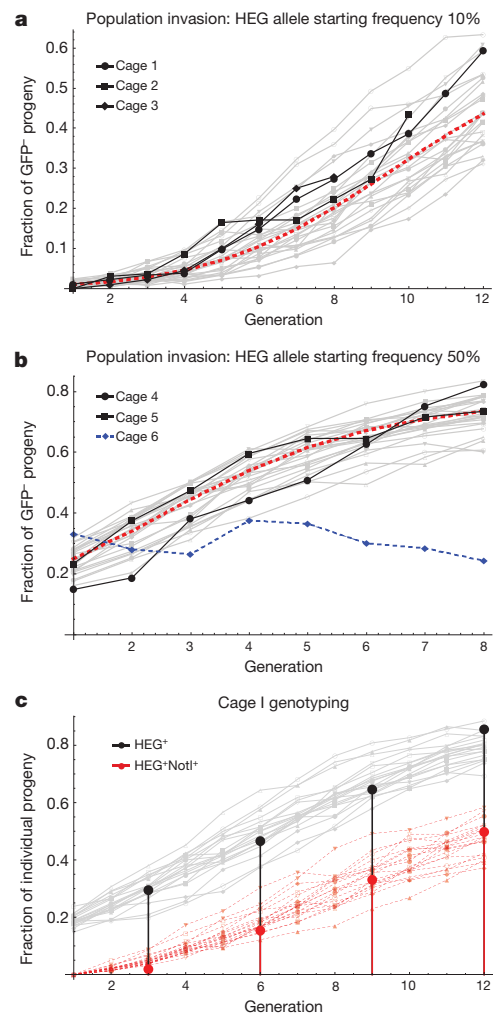


Figure 2 | HEG invasion in mosquito cage populations. **a, b,** Temporal dynamics of GFP^- mosquitoes in populations in which the HEG Donor allele was seeded at a frequency of 10% (**a**) or 50% (**b**) into a background of GFP^+ mosquitoes carrying the HEG Target allele. The experimental points (black) are overlaid onto predicted dynamics derived from a deterministic population genetic model (dashed red line) and from 20 iterations of a stochastic model (grey lines). The dynamics of a cage population in which the HEG Donor and Control alleles were combined at a frequency of 50% is also shown (dashed blue line). **c,** Molecular genotyping performed on individuals randomly collected from cage 1 at generations 3, 6, 9 and 12 using a set of PCR primers that specifically amplifies the HEG cassette (Primer set 1b, Supplementary Fig. 2). Presence of the NotI marker was determined by *in vitro* digestion of PCR products using NotI. The graph shows the fraction of mosquitoes carrying the HEG (black) and the fraction carrying the HEG and the NotI marker on the same chromosome (red) overlaid onto predictions from 20 stochastic simulations (grey lines and dashed red lines, respectively).

nuclear genomes of metazoans; our results indicate that this absence is not because homing cannot occur, and instead supports alternative explanations such as that the segregated germline of animals prevents the horizontal transmission amongst species that these selfish genes need to persist over long evolutionary timespans¹⁷. The transmission dynamics of HEGs in cage populations provide the first evidence of the potential of these genetic elements to serve as synthetic gene drive systems in insect pests and add a promising candidate to those under development^{18,19}.

The sequence-specific activity of HEGs could be exploited to develop vector control strategies aimed at either disrupting the mosquito genes that contribute to its vectorial capacity or introducing at selected chromosomal locations novel genes that impair the mosquito's ability to function as vector for malaria⁹. Any use of HEGs in natural

A. gambiae populations will depend on the ability to re-engineer their specificity^{20–23} towards native mosquito sequences. We identified in the *A. gambiae* genome within intergenic regions of the left (2L) and right arms (2R) of chromosome 2 two sequences that show similarities to the recognition sites of the two HEGs I-AniI and I-CreI that have previously been shown to be amenable to re-engineering to target novel human and plant sequences^{23–29}. A previously described HEG engineering strategy was then used to generate an I-AniI variant to selectively cleave the 2L site, and a variant of monomerized I-CreI (termed mCre³⁰) to cleave the 2R site selectively (Supplementary Fig. 6). The change in specificity of these enzymes demonstrates that HEGs can be designed to recognize new mosquito sequences and opens the possibility to investigate the biology of HEGs in wild-type mosquito populations. Although technical hurdles in HEG engineering technology must still be addressed to reach the flexibility required to target specific mosquito genes essential for viability or disease transmission, our results suggest how these genetic elements could overcome a major scientific roadblock in developing genetic control measures targeting species like the main vector of human malaria: the genetic manipulation of entire field populations starting from a few laboratory individuals.

METHODS SUMMARY

The generation of transgenic lines and population cage experiments are described in Methods. To monitor homing mosquitoes were subjected to fluorescent microscopy at the larval stage to detect the presence of the marker genes or subjected to PCR to detect the presence of the HEG gene at the adult stage.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 September 2010; accepted 16 February 2011.

Published online 20 April 2011.

- Curtis, C. F. Possible use of translocations to fix desirable genes in insect pest populations. *Nature* **218**, 368–369 (1968).
- Hamilton, W. D. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science* **156**, 477–488 (1967).
- Alphey, L. *et al.* Malaria control with genetically manipulated insect vectors. *Science* **298**, 119–121 (2002).
- Corby-Harris, V. *et al.* Activation of Akt signaling reduces the prevalence and intensity of malaria parasite infection and lifespan in *Anopheles stephensi* mosquitoes. *PLoS Pathog.* **6**, e1001003 (2010).
- Ito, J., Ghosh, A., Moreira, L. A., Wimmer, E. A. & Jacobs-Lorena, M. Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature* **417**, 452–455 (2002).
- Moreira, L. A. *et al.* Bee venom phospholipase inhibits malaria parasite development in transgenic mosquitoes. *J. Biol. Chem.* **277**, 40839–40843 (2002).
- Li, F., Patra, K. P. & Vinetz, J. M. An anti-chitinase malaria transmission-blocking single-chain antibody as an effector molecule for creating a *Plasmodium falciparum*-refractory mosquito. *J. Infect. Dis.* **192**, 878–887 (2005).
- Sinkins, S. P. & Gould, F. Gene drive systems for insect disease vectors. *Nature Rev. Genet.* **7**, 427–435 (2006).
- Burt, A. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. R. Soc. Lond. B* **270**, 921–928 (2003).
- Catteruccia, F., Benton, J. P. & Crisanti, A. An *Anopheles* transgenic sexing strain for vector control. *Nature Biotechnol.* **23**, 1414–1417 (2005).
- Jacquier, A. & Dujon, B. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* **41**, 383–394 (1985).
- Bellaiche, Y., Mogila, V. & Perrimon, N. I-SceI endonuclease, a new tool for studying DNA double-strand break repair mechanisms in *Drosophila*. *Genetics* **152**, 1037–1044 (1999).
- Windbichler, N. *et al.* Homing endonuclease mediated gene targeting in *Anopheles gambiae* cells and embryos. *Nucleic Acids Res.* **35**, 5922–5933 (2007).
- Stoddard, B. L. Homing endonuclease structure and function. *Q. Rev. Biophys.* **38**, 49–95 (2005).
- Goddard, M. R., Greig, D. & Burt, A. Outcrossed sex allows a selfish gene to invade yeast populations. *Proc. R. Soc. Lond. B* **268**, 2537–2542 (2001).
- Meredith, J. M. *et al.* Site-specific integration and expression of an anti-malarial gene in transgenic *Anopheles gambiae* significantly reduces *Plasmodium* infections. *PLoS ONE* **6**, e14587 (2011).
- Burt, A. & Koufopanou, V. Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr. Opin. Genet. Dev.* **14**, 609–615 (2004).
- Chen, C. H. *et al.* A synthetic maternal-effect selfish genetic element drives population replacement in *Drosophila*. *Science* **316**, 597–600 (2007).
- McMeniman, C. J. *et al.* Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*. *Science* **323**, 141–144 (2009).
- Ashworth, J. *et al.* Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* **441**, 656–659 (2006).
- Jarjour, J. *et al.* High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display. *Nucleic Acids Res.* **37**, 6871–6880 (2009).
- Ashworth, J. *et al.* Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res.* **38**, 5601–5608 (2010).
- Thyme, S. B. *et al.* Exploitation of binding energy for catalysis and design. *Nature* **461**, 1300–1304 (2009).
- Gao, H. *et al.* Heritable targeted mutagenesis in maize using a designed endonuclease. *Plant J.* **61**, 176–187 (2010).
- Grizot, S. *et al.* Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res.* **37**, 5405–5419 (2009).
- Munoz, I. G. *et al.* Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. *Nucleic Acids Res.* **39**, 729–743 (2010).
- Redondo, P. *et al.* Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature* **456**, 107–111 (2008).
- Arnould, S. *et al.* Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J. Mol. Biol.* **371**, 49–65 (2007).
- Rosen, L. E. *et al.* Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Res.* **34**, 4791–4800 (2006).
- Li, H., Pellenz, S., Ulge, U., Stoddard, B. L. & Monnat, R. J. Jr. Generation of single-chain LAGLIDADG homing endonucleases from native homodimeric precursor proteins. *Nucleic Acids Res.* **37**, 1650–1662 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Ashburner, S. Russell, D. Huen and S. Chan for comments, assistance and for plasmids. We thank M. P. Calos for providing the pET11phiC31polyA plasmid. We thank M. J. Fraser Jr for providing the pBSII-IFP2-orf plasmid. We thank J. Meredith and P. Eggleston for providing the docking strain. We thank A. Hall, T. Nolan, K. Magnusson, D. Rogers and S. Fuchs for assistance. We thank S. Arshiya Quadri and M. Szeto for experimental support and the members of the laboratories of D. Baker, R. Monnat, A. Scharenberg and B. Stoddard for their collective support of HEG engineering. A. F. M. Hackmann provided graphics support. Funded by a grant from the Foundation for the National Institutes of Health through the Vector-Based Control of Transmission: Discovery Research (VCTR) program of the Grand Challenges in Global Health initiative and by NIH R1L1 awards GM084433 to D.B. and CA133831 to R.J.M.

Author Contributions N.W. designed the experiments. N.W., M.M. and P.A.P. performed the experiments. N.W. and P.A.P. generated the transgenic lines. M.M. maintained mosquito populations. N.W. analysed the data. A.B. and N.W. generated the population dynamic models. A.C. and A.B. inspired the work and wrote the paper together with N.W. HEG redesign and target site cleavage analyses were performed by S.B.T., H.L., U.Y.U. (contributed equally) and B.T.H. with guidance from D.B. and R.J.M. All authors read and approved the final manuscript.

Author Information The plasmids pHome-T and pHome-D have been deposited to GenBank under the accession numbers HQ159398 and HQ159399. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.C. (adrcrisanti@imperial.ac.uk).

METHODS

Development of transgenic lines. The pHome-R, pHome-T (GenBank HQ159398) and pHome-D (GenBank HQ159399) plasmids are derived from the same parent backbone and are identical apart from differences explained below. All three plasmids contain the DsRed fluorescent protein (RFP) reporter gene driven by the *Drosophila* Actin5C promoter as well as piggyBac inverted repeats for transposase mediated integration and an AttB sequence for site specific integration using the ϕ C31 integrase. They also contain a 3 \times P3 (artificial promoter element binding 3 Pax-6 homodimers³¹) driven GFP marker gene that is modified the following way: the pHome-T construct contains the 18-base pairs I-SceI target site (The *A. gambiae* genome does not contain an I-SceI site) within the open reading frame of the GFP reporter gene (Supplementary Fig. 2b). The GFP coding sequence (CDS) containing the I-SceI recognition sequence remains functional but can be inactivated by cleavage followed by certain types of non-homologous end joining (NHEJ) DNA repair events (in particular when followed by a frameshift as the GFP N terminus is generally tolerant to amino acid changes) or homing. In addition the pHome-T construct carries a NotI recognition site immediately downstream of the GFP open reading frame (ORF) that is not present in the other two plasmids. The pHome-R construct also contains an I-SceI site within the GFP gene. In this case the GFP ORF of this plasmid is out of frame but cleavage and NHEJ followed by a frameshift is expected to restore GFP expression (Supplementary Fig. 1b). Finally the pHome-D construct contains the HEG expression cassette which consists of the *Anopheles* β 2-tubulin promoter¹⁰ and terminator flanking the I-SceI ORF, which contains an SV40 nuclear localization signal. This cassette is located within two I-SceI half sites disrupting the GFP gene at the same position as the I-SceI recognition site in the Reporter and Target vectors. This setup resembles the natural occurrence of HEGs and their target sites as both the HEG expression cassette and the HEG target site are flanked by identical homologous regions. Transgenic lines were developed as previously described^{10,32}. Briefly, *A. gambiae* embryos of the ϕ C31 integrase docking line¹⁶ (or wild-type embryos) were injected using a Femtojet Express injector and sterile Femtotips (Eppendorf) with a mixture of 0.2 mg ml⁻¹ of plasmid and 0.8 mg ml⁻¹ of ϕ C31 integrase or piggyback helper RNA, respectively. The 5'-capped helper RNAs were produced using the mMESSAGEmachine kit (Ambion) from linearized vectors pBSII-IFP2-orf (transposase) and pET11phiC31polyA (integrase). The hatched larval survivors were screened for transient expression of either the 3 \times P3-GFP if present or the Actin5C-RFP marker. In the presence of the 3 \times P3-GFP marker (pHome-T) only transients were grown up and crossed to wild-type mosquitoes whereas in the case of Actin5C-RFP (pHome-R, pHome-D) all survivors were crossed as this promoter drives no expression in the most posterior segments of the larvae where, due to the way embryos are injected, most or all of the transient fluorescence is usually observed. The progeny of these crosses were analysed for fluorescence to identify transgenic individuals. We have previously shown that founder effects and inbreeding can be determinants of the fitness of transgenic mosquitoes³³. To minimize these effects the progeny of each transgenic founder was backcrossed to wild-type mosquitoes for at least three generations before homozygote strains were established. Transgenic mosquitoes at different developmental stages were analysed on a Nikon inverted microscope (Eclipse TE200) to detect GFP, RFP and CFP expression. Digital images were captured on a Nikon inverted microscope (Eclipse TE200) with an attached Nikon DXM1200 digital camera.

PCR and restriction analysis. PCRs (Phusion HF polymerase, Finnzymes) were performed on genomic DNA (Wizard Genomic DNA purification kit, Promega) prepared from single transgenic adult mosquitoes. We extracted DNA from single hemizygous virgin female or male mosquitoes in the mating experiments. Each hemizygous offspring analysed allows the scoring of a single chromosome from the double transgenic parent. We extracted DNA from single virgin female adults in the case of the population cage experiments and the PCR assay was performed on 60 to 96 randomly chosen individuals per generation. The following primers as shown in Supplementary Fig. 1 were used:

Primer set 1a, forward 5'-TGGAAATGAGAAGTAGGTGCATCTGCA-3', reverse 5'-GGAATAAGGGCGACACGGAATGTTG-3'; primer set 1b, forward 5'-TGTGACAGTGGAAATGAGAAGTAGGTGC-3', reverse 5'-TCTCAACGTAGTCCAAAGCATCAA-3'; primer set 2, forward 5'-GCGATGACGAGCTTGTGGTG-3', reverse 5'-CGTGACAGGCTTTGATAACTCT-3'; primer set 3, forward 5'-CTCTCGCTCTCAAGTCGCGTTCA-3', reverse 5'-TGCAGATGCACCTACTTCTCATTTC-3'; primer set 4, forward 5'-ATCGTGATAGGTGCCTCACTGA-3', reverse 5'-CTCATGTAAACAGTTCA TAGTTCTCGC-3'.

In vitro digestions using NotI (Roche) and I-SceI (NEB) were performed according to manufacturer's recommendations. We used primer set 1a on hemizygous progeny of Donor/Target and Donor/Reporter crosses. Primer set 1b was used in Donor/Ectopic Target crosses and all population cage experiments.

Population dynamic modelling. To model the cage populations we assumed that the two starting alleles, Target (*T*; GFP⁺, HEG⁻, NotI⁺), and Donor (*D*; GFP⁻, HEG⁺, NotI⁻), give rise to three classes of novel alleles: (1) *D*^N, products of homing that retain the NotI site and can themselves act as donors (GFP⁻, HEG⁺, NotI⁺); (2) *M*⁺, products of misrepair (for example, homologous repair from a different template, or non-homologous end-joining) that are GFP⁺ (that is, GFP⁺, HEG⁻, NotI^{+/−}); and (3) *M*⁻, products of misrepair that are GFP⁻ (GFP⁻, HEG⁻, NotI^{+/−}). All three novel alleles are resistant to further cleavage, and the two products of misrepair are not able to home. In the germlines of male *D/T* or *D^N/T* mosquitoes the *T* allele is cleaved with probability *c*. In *D^N/T* males these cleaved alleles are then converted into a *D^N* allele with probability *h*, into an *M*⁺ allele with probability $(1-h)r$, or into an *M*⁻ allele with probability $(1-h)(1-r)$, where *h* is the rate of canonical homologous repair in males and *r* is the probability that other forms of repair maintain GFP expression. In *D/T* males, cleaved *T* alleles are converted into *D^N* alleles with probability *hn* and into *D* alleles with probability $h(1-n)$, where *n* is the probability that new products of homing retain the NotI site, and $1-n$ is the probability that it is lost by co-conversion. The overall net homing rate in this model is $e_m = ch$.

Estimates of *c*, *h*, *n* and *r* were derived from the Donor/Target experiments as follows.

(1) The fraction of GFP⁺ progeny from *D/T* males is 0.14. PCR and sequencing of 36 such individuals showed that 7 had the intact Target sequence (19.4%). The proportion of gametes with uncleaved *T* alleles is therefore $0.14 \times 0.194 = 0.0272$. Because only half the gametes should carry the *T* allele or its descendants (the other half being derived from the Donor allele), the cleavage rate in males is $c = 1 - 2 \times 0.0272 = 0.95$ (that is, 95%).

(2) The fraction of progeny from *D/T* males that are GFP⁻ RFP⁺ CFP⁺ is 0.803. PCR analysis of 156 such individuals showed that 152 of them were HEG⁺ (97.4%). The proportion of gametes that are HEG⁺ is thus $0.803 \times 0.974 = 0.782$. Because 50% of gametes are expected to be HEG⁺ in the absence of homing, the estimated homing rate is $e = 2(0.782 - 0.5) = 0.564$. Because the homing rate is equal to the product of the cleavage rate and the rate of canonical homologous repair, the latter is estimated to be $h = 0.564/0.95 = 0.60$ (that is, 60%).

(3) Of the 152 HEG⁺ progeny described above, 25 were also NotI⁺. Therefore, the overall fraction of HEG⁺ NotI⁺ progeny is $0.803 \times 25/156 = 0.129$. The fraction of gametes with a newly acquired HEG is $e/2 = 0.281$. Therefore, the probability that homing leads to retention of the NotI site is $n = 0.129/0.282 = 0.45$.

(4) The fraction of progeny from *D/T* males that were GFP⁺ and that molecular analysis showed had been cleaved and misrepaired was 0.112. The total fraction of gametes with misrepaired alleles (that is, cleaved and not subject to canonical homing) is $c(1-h)/2 = 0.19$. Therefore the probability that misrepaired alleles remain GFP⁺ is $r = 0.112/0.19 = 0.58$.

Populations start as a mixture of *T/T* and *D/D* homozygotes. *T/D* heterozygous males produce sperm carrying alleles *T*, *D*, *M*⁻, *M*⁺ or *D^N* with probabilities $(1 \times c)/2$, $1/2 + ch(1 \times m)/2$, $c(1-h)(1-r)/2$, $c(1-h)r/2$, and $chm/2$, respectively, and *T/D^N* heterozygous males produce sperm carrying these alleles with probabilities $(1-c)/2$, 0, $c(1-h)(1-r)/2$, $c(1-h)r/2$, and $1/2 + ch/2$, respectively. All other male genotypes and all female genotypes produce gametes in Mendelian proportions. All genotypes have equal survival and fertility: each female mates with a single male, chosen randomly (with replacement), and each offspring is from a randomly chosen mated female (with replacement). Simulations were generated in the Mathematica software suite 7 (Wolfram Research).

Defined crosses and population cage experiments. Crosses were carried out using 25 males and 25 virgin females. A total of 1,996 (Donor/Target), 3,582 (Donor/Reporter) and 720 (Donor/Ectopic Target) offspring were analysed in at least three independent experiments for green (GFP), blue (CFP) and red (RFP) fluorescence. Independently reared cage populations were established and the I-SceI containing allele was seeded at a frequency of 10% or 50%. To achieve a frequency of 10% the cage (BugDorm-1, Megaview) contained a population of 540 homozygote mosquitoes carrying the I-SceI Target construct (270 males and 270 females). In addition the population cage also contained 60 homozygote mosquitoes carrying the I-SceI Donor construct (30 males and 30 females). Each generation mosquitoes were allowed to mate for 5–7 days, and then fed on 2–3 mice to ensure that all females were blood-fed. Larvae were allowed to hatch from the eggs and reared until the L3–L4 stage, at which point a random set of at least 300 was screened for the presence of the fluorescent markers. At the pupal stage mosquitoes were separated according to sex and males and females were allowed to emerge separately for each cage population. After at least 48 h, 300 male and 300 female mosquitoes of each population were added to a fresh cage to establish the next generation.

Identification and cloning of *Anopheles gambiae* genomic target sites. Genomic targets for engineered I-AniI and mCreI protein variants were identified

by searching the *Anopheles gambiae* genome with PSSMs (positional specific scoring or search matrices) constructed for each HEG protein. The I-AniI PSSM was constructed from cleavage degeneracy data, computational design results and selected variants isolated by using a modified bacterial selection system^{34–36} (additional unpublished results). The mCreI PSSM was constructed from cleavage degeneracy data and the results of a comprehensive computational design analysis of I-CreI design data^{37,38} (additional unpublished results). A predicted cleavage-sensitive *Anopheles gambiae* chromosomal target site for an engineered I-AniI protein that combined two unpublished protein variants was identified on chromosome 2L (Agam 2L –3C/+5G site: reverse strand nucleotides 26449203–26449184). A comparable, predicted cleavage-sensitive chromosomal target site for a –5C mCreI design was identified on chromosome 2R (Agam 2R –5C2 site: forward strand nucleotides 33439283–33439302). Both target sites, together with 15-bp of flanking *Anopheles* chromosomal sequence on each side, were synthesized as pairs of complementary oligonucleotides that were annealed and ligated into the NheI/SacII sites of the bacterial plasmid vector *pCcdB* to facilitate cleavage analyses³⁶. A native I-AniI site previously cloned into pBluescript³⁵ and a native I-CreI target site cloned into *pCcdB* were used as positive control sites.

Engineering of *Anopheles* target site-specific I-AniI and mCreI variants. An engineered I-AniI variant predicted to cleave the Agam 2L –3C/+5G chromosomal target site was generated by combining two previous engineered variants for the –3C and +5G base pair positions in the M4 variant of I-AniI (previously described Y2 variant + two additional residue substitutions, F91I and S92T)^{35,39}. The –3C variant was based on a previously published –3C computational design³⁵ that was further improved by bacterial selection³⁶. The residue substitutions in this I-AniI variant were Y18W, E35K, and substitution of the four residue loop PDGM for the native 7-residue loop between I-AniI positions K60 and M66. The +5G variant was identified in a bacterial selection and contained a D168Q substitution. These modifications were combined and incorporated into the open reading frame of the M4 Y2 variant of I-AniI^{35,39}.

Variants of mCreI specific for the *Anopheles* 2R –5C2 mCreI chromosomal target site were generated using RosettaDesign (RD), a macromolecular modelling and design suite⁴⁰. In brief, the –5G>C base change in the *Anopheles* 2R mCreI target site was modelled, and amino acid residues in close proximity to the –5 position were allowed to mutate *in silico* to accommodate the new –5C design target base pair. Amino acid conformations and associated hydration patterns that improved the energy of mCreI –5C target site complex were accepted more often during design runs, and converged to identify energetically favourable amino acid substitutions predicted to be specific for the –5C design target. The resulting residue substitutions, I24K and R68T, were incorporated into the open reading frame of mCreI by PCR-mediated mutagenesis³⁸. The resulting engineered I-AniI and mCreI variant proteins and native control proteins were expressed in *E. coli* and purified as previously described for *in vitro* cleavage analyses^{35,38,41}. Bacterial selection³⁶ has already been used to improve Y2 I-AniI and to generate

mCreI^{30,35,39}. Thus it should be possible to use sequential positive and negative bacterial selection to rapidly improve further the cleavage efficiency of both engineered proteins on their *Anopheles* chromosomal target sites, and suppress residual cleavage activity on their native target sites if required.

In vitro cleavage assays. pCcdB vector DNA containing the *Anopheles* 2L –3C/+5G I-AniI chromosomal target site was linearized with XbaI, and pBluescript plasmid DNA containing the native I-AniI target site with ScaI, before the cleavage analyses. Cleavage reactions (10 µl final volume) were performed in digest buffer (170 mM KCl, 10 mM MgCl₂, 20 mM Tris pH 9.0) containing 5 nmol linearized target site plasmid and serial twofold dilutions of purified enzyme ranging from 800 nM to 12.5 nM. Reactions were incubated at 37 °C for 0.5 h, then stopped by the addition of stop buffer (200 mM EDTA, 30% glycerol, bromophenol blue) before agarose gel electrophoresis to separate substrate and cleavage products.

pCcdB vector DNA containing the *Anopheles* 2R –5C2 mCreI chromosomal target site was linearized by NcoI digestion before cleavage analyses. Cleavage reactions (10 µl final volume) were performed in digest buffer (10 mM MgCl₂, 20 mM Tris pH 8.0) containing 10 nmol linearized target site plasmid and serial twofold dilutions of purified enzyme ranging from 320 nM to 10 nM. Reactions were incubated at 37 °C for 1 h, then stopped by the addition of stop buffer (0.5% SDS and bromophenol blue) before agarose gel electrophoresis.

31. Sheng, G., Thouvenot, E., Schmucker, D., Wilson, D. S. & Desplan, C. Direct regulation of *rhodopsin 1* by *Pax-6/eyeless* in *Drosophila*: evidence for a conserved function in photoreceptors. *Genes Dev.* **11**, 1122–1131 (1997).
32. Lobo, N. F., Clayton, J. R., Fraser, M. J., Kafatos, F. C. & Collins, F. H. High efficiency germ-line transformation of mosquitoes. *Nature Protocols* **1**, 1312–1317 (2006).
33. Catteruccia, F., Godfray, H. C. & Crisanti, A. Impact of genetic manipulation on the fitness of *Anopheles stephensi* mosquitoes. *Science* **299**, 1225–1227 (2003).
34. Scalley-Kim, M., McConnell-Smith, A. & Stoddard, B. L. Coevolution of a homing endonuclease and its host target sequence. *J. Mol. Biol.* **372**, 1305–1319 (2007).
35. Thyme, S. B. et al. Exploitation of binding energy for catalysis and design. *Nature* **461**, 1300–1304 (2009).
36. Doyon, J. B., Pattanayak, V., Meyer, C. B. & Liu, D. R. Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J. Am. Chem. Soc.* **128**, 2477–2484 (2006).
37. Argast, G. M., Stephens, K. M., Emond, M. J. & Monnat, R. J. Jr. I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential *in vitro* enrichment. *J. Mol. Biol.* **280**, 345–353 (1998).
38. Ulge, U. Y., Baker, D. A. & Monnat, R. J. Comprehensive computational design of mCreI homing endonuclease cleavage specificity for genome engineering. *Nucleic Acids Res.* doi:10.1093/nar/gkr022 (1 February 2011).
39. McConnell Smith, A. et al. Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc. Natl Acad. Sci. USA* **106**, 5099–5104 (2009).
40. Das, R. & Baker, D. Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* **77**, 363–382 (2008).
41. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).

Metabolite-enabled eradication of bacterial persisters by aminoglycosides

Kyle R. Allison¹, Mark P. Brynildsen^{1†} & James J. Collins^{1,2,3}

Bacterial persistence is a state in which a sub-population of dormant cells, or ‘persisters’, tolerates antibiotic treatment^{1–4}. Bacterial persisters have been implicated in biofilms and in chronic and recurrent infections^{5–7}. Despite this clinical relevance, there are currently no viable means for eradicating persisters. Here we show that specific metabolic stimuli enable the killing of both Gram-negative (*Escherichia coli*) and Gram-positive (*Staphylococcus aureus*) persisters with aminoglycosides. This potentiation is aminoglycoside-specific, it does not rely on growth resumption and it is effective in both aerobic and anaerobic conditions. It proceeds by the generation of a proton-motive force which facilitates aminoglycoside uptake. Our results demonstrate that persisters, although dormant, are primed for metabolite uptake, central metabolism and respiration. We show that aminoglycosides can be used in combination with specific metabolites to treat *E. coli* and *S. aureus* biofilms. Furthermore, we demonstrate that this approach can improve the treatment of chronic infections in a mouse urinary tract infection model. This work establishes a strategy for eradicating bacterial persisters that is based on metabolism, and highlights the importance of the metabolic environment to antibiotic treatment.

Translation occurs at a reduced rate in bacterial persisters^{2,8}, indicating that they should be susceptible to aminoglycosides, which are ribosome-targeting bactericidal antibiotics^{9–13}. However, aminoglycosides have weak activity against dormant bacteria, despite continued translation in these cells^{14,15}. Given the dormancy of persisters and the known energy requirement for aminoglycoside activity¹⁶, we reasoned that metabolic stimulation might potentiate aminoglycosides against bacterial persisters.

To test this, we screened metabolites for their ability to potentiate aminoglycosides against *E. coli* persisters. We selected carbon sources to maximize coverage of glycolysis, the pentose-phosphate pathway (PPP) and the Entner-Doudoroff pathway (EDP) (Fig. 1a, b). Persisters were isolated (see Methods), re-suspended in minimal media supplemented with individual metabolites and then treated for 2 h with the aminoglycoside gentamicin.

We found that gentamicin was significantly potentiated against persisters by specific metabolic stimuli (Fig. 1a, b). Metabolites that enter upper glycolysis (glucose, mannitol and fructose), as well as pyruvate, induced rapid killing of persisters by gentamicin, reducing persister viability by three orders of magnitude. In contrast, metabolites that enter lower glycolysis (except for pyruvate) caused little potentiation. Metabolites entering metabolism via the PPP or EDP (arabinose, ribose and gluconate) also showed low potentiation. No killing was observed in the control, demonstrating that the treated cells were persistent to gentamicin in the absence of an added metabolite. We verified that the metabolite-enabled eradication of persisters was general to the aminoglycoside class of antibiotics by testing kanamycin and streptomycin (Supplementary Fig. 2).

We considered that the potentiating metabolites might cause persisters to revert to normally growing cells, which would render them

susceptible to quinolone (DNA-damaging) and β -lactam (cell-wall-inhibiting) antibiotics. To test this, we treated persisters in the presence and absence of mannitol with a member of each of the three major classes of bactericidal antibiotics: aminoglycosides, quinolones and β -lactams. As seen in the metabolite screen, gentamicin rapidly eliminated metabolically stimulated persisters (Fig. 1c). However, neither the β -lactam ampicillin nor the quinolone ofloxacin showed appreciable killing of persisters in either the presence or the absence of mannitol. This shows that potentiation is aminoglycoside-specific and that cells were persistent to quinolones and β -lactams. It further indicates that metabolic stimuli under these conditions do not rapidly revert persisters to a growth state in which cell-wall synthesis and DNA synthesis are active. To explore this further, we tested the growth of persisters on the metabolites used for aminoglycoside potentiation and observed negligible growth of persisters 8 h after metabolite addition (Supplementary Figs 3 and 4). Taken together, these data indicate that the metabolic stimuli bolster a process that is specific to aminoglycosides and do not cause persisters to revert to normally growing cells.

Given that aminoglycoside uptake is energy-dependent¹⁶, we investigated whether the metabolic stimuli screened here could increase aminoglycoside uptake. We measured uptake by fluorescently labeling gentamicin with Texas red and analysing cells by fluorescence-activated cell sorting (FACS). Cells were pre-incubated with metabolites for 30 min before a 5-min treatment with gentamicin-Texas red (Fig. 1d and Supplementary Fig. 10). Metabolites that induced substantial killing by aminoglycosides also induced high levels of aminoglycoside uptake, implying that the increased uptake was responsible for the killing. Furthermore, metabolites that caused low potentiation did not significantly increase aminoglycoside uptake.

The requirement of proton-motive force (PMF) for aminoglycoside uptake in exponentially growing bacteria has been studied extensively¹⁶. Although the complete mechanism of aminoglycoside uptake is unclear, it is known that a threshold PMF is required. We reasoned that, although metabolic stimuli do not rapidly stimulate the growth of persisters, they may promote PMF, thereby facilitating the uptake of aminoglycosides and subsequent bacterial killing. To test this hypothesis, we pre-incubated persisters with the proton ionophore carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP), which inhibits PMF, before treating them with metabolites in conjunction with gentamicin. Treatment with CCCP was found to abolish aminoglycoside potentiation by all of the carbon sources, demonstrating that PMF, induced by metabolites, is required for persister elimination (Fig. 2a and Supplementary Fig. 12). We next verified that the requirement for PMF was due to aminoglycoside uptake. We pre-incubated samples with CCCP and performed gentamicin-Texas red uptake experiments, showing that inhibiting PMF suppressed the metabolite-induced uptake of aminoglycoside (Fig. 2b and Supplementary Fig. 13). Furthermore, using the membrane stain 3,3'-diethyloxycarbocyanine iodide (DiOC₂(3)), we verified that metabolites that induced aminoglycoside uptake and bacterial killing were also the ones that induced an elevated

¹Howard Hughes Medical Institute, Department of Biomedical Engineering, Center for Biodynamics and Center for Advanced Biotechnology, Boston University, Boston, Massachusetts 02215, USA. ²Boston University School of Medicine, 715 Albany Street, Boston, Massachusetts 02118, USA. ³Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02118, USA.

[†]Present address: Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, USA.

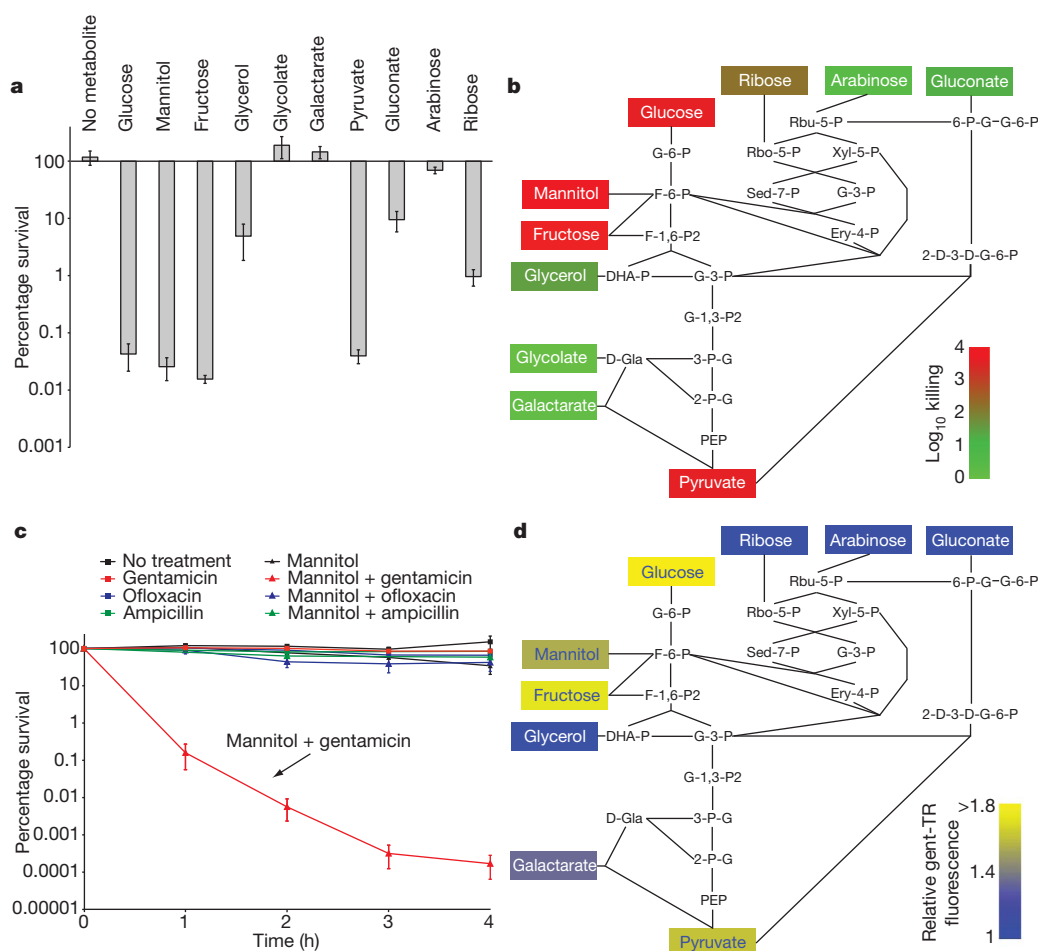


Figure 1 | Specific metabolites enable killing of *E. coli* persisters by aminoglycosides. **a**, Percentage survival of persisters after 2-h treatment with gentamicin and metabolites. **b**, Efficiency of metabolite-enabled elimination of persisters, superimposed on a metabolic network. Metabolites are colour-coded according to the aminoglycoside killing which they induce (\log_{10} killing). The following metabolic intermediates are depicted: glucose-6-phosphate (G-6-P), fructose-6-phosphate (F-6-P), fructose-1,6-bisphosphate (F-1,6-P2), dihydroxyacetone phosphate (DHAP), glyceraldehyde-3-phosphate (G-3-P), 1,3-bisphospho-glycerate (G-1,3-P2), 3-phospho-glycerate (3-P-G), 2-phospho-glycerate (2-P-G), phosphoenolpyruvate (PEP), D-glycerate

(D-Gla), ribulose-5-phosphate (Rbu-5-P), ribose-5-phosphate (Rbo-5-P), xylulose-5-phosphate (Xyl-5-P), glyceraldehyde-3-phosphate (G-3-P), erythrose-4-phosphate (Ery-4-P), sedoheptulose-7-phosphate (Sed-7-P), 6-phospho-gluconate (6-P-G) and 2-dehydro-3-deoxy-gluconate-6-phosphate (2-D-3-D-G-6-P). **c**, Percentage survival of persisters after treatment with mannitol and different classes of antibiotics. **d**, Metabolite-enabled uptake of gentamicin-Texas red (gent-TR) by stationary-phase cells, superimposed on a metabolic network (see also Supplementary Fig. 10). Metabolites are colour-coded according to the relative uptake that they induce. Means \pm s.e.m. are presented in **a** and **c** ($n \geq 3$).

PMF (Supplementary Figs 14 and 15). These results demonstrate that specific metabolites induce PMF in persisters, thereby facilitating aminoglycoside uptake and bacterial killing.

From these results, we hypothesized that aerobic respiration is primed in persisters and that it facilitates the metabolic potentiation of aminoglycosides. We tested this using genetic knockout strains inactivated for each of the *E. coli* cytochrome quinol oxidases (bo, $\Delta cyoA$; bd-I, $\Delta cydB$; bd-II, $\Delta appB$). We also used potassium cyanide to inhibit all cytochromes simultaneously. Wild-type persisters, with and without potassium cyanide, and enzymatically inactivated persisters were treated for 2 h with gentamicin plus metabolites (Fig. 2c and Supplementary Fig. 16). Treatment with potassium cyanide abolished bacterial killing, consistent with studies of rapidly growing bacteria¹⁷, demonstrating the necessity of aerobic respiration for elimination of persisters by aminoglycosides under these conditions. The $\Delta cydB$ mutation, which abolishes activity of the microaerobic cytochrome bd-I^{18,19}, suppressed killing by more than two orders of magnitude, possibly owing to the use of cytochrome bd-I in oxygen-depleted and alkaline stationary-phase cultures. Neither $\Delta cyoA$ nor $\Delta appB$ showed a significant effect on bacterial killing. Although we found that aerobic respiration was required for eradication of bacteria in aerated conditions, we

also found that metabolite-enabled eradication occurs anaerobically in conditions that support PMF (Supplementary Figs 17 and 18).

Because aerobic respiration in *E. coli* is driven by NADH oxidation, we investigated the role of NADH utilization in this phenotype. Persister cells inactivated for NADH dehydrogenase I ($\Delta nuoI$), NADH dehydrogenase II (Δndh) and both NADH dehydrogenases ($\Delta ndh\Delta nuoI$) were treated for 2 h with gentamicin plus metabolites (Fig. 2d and Supplementary Fig. 19). We found that NADH dehydrogenase activity was important for this phenotype because gentamicin activity against the $\Delta ndh\Delta nuoI$ strain was not potentiated by mannitol, fructose or pyruvate, although there was slight potentiation by glucose (Supplementary Fig. 19a). Given that NADH drives electron transport, this requirement for NADH is not surprising, although we found that it is not essential for killing under all conditions (Supplementary Figs 18 and 20). Both the Δndh and $\Delta nuoI$ deletions suppressed killing but the $\Delta nuoI$ strain was the more resistant of the two, possibly reflecting the direct contribution of NADH dehydrogenase I to PMF. Using a series of genetic knockouts, we further determined that the enzyme pyruvate dehydrogenase was necessary for the observed phenotype, owing to its generation of NADH, whereas the PPP, EDP and TCA cycle were not necessary (Supplementary Figs 21–24).

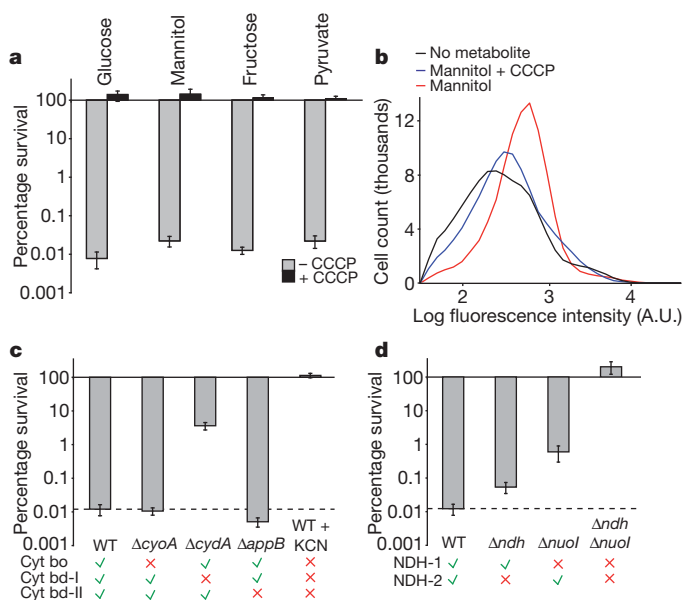


Figure 2 | Metabolite-enabled aminoglycoside uptake and bacterial killing requires PMF produced by the oxidative electron transport chain.

a, Percentage survival of persisters after treatment with gentamicin plus uptake-potentiating metabolites with CCCP (dark grey bars) and without CCCP (light grey bars). **b**, Representative measurement of gentamicin-Texas red uptake by stationary-phase cells after incubation with no metabolite, mannitol or mannitol and CCCP (see also Supplementary Fig. 13). A.U., arbitrary units. **c**, Percentage survival of persisters in wild-type (WT) and cytochrome-inactivated strains after treatment with gentamicin plus mannitol (see also Supplementary Fig. 16). The presence (green ticks) or absence (red crosses) of functional complexes is indicated below test conditions. **d**, Percentage survival of persisters in wild-type and NADH-dehydrogenase-inactivated strains after treatment with gentamicin plus mannitol (see also Supplementary Fig. 19). The presence or absence of functional complexes is indicated as in **c**. Means \pm s.e.m. are presented throughout ($n \geq 3$).

These results show that persisters are primed for specific biochemical processes that allow PMF induction, including central metabolism. However, in the timescales examined here, this resumption of central

metabolism and respiration was not sufficient to support other processes necessary for cellular growth, such as cell-wall biogenesis and DNA replication. Thus, persisters treated with specific metabolites seem to be in an energized but non-dividing state that facilitates their elimination by aminoglycosides. On the basis of these findings, we propose the following mechanism for metabolite-enabled eradication of persisters by aminoglycosides (Fig. 3a). Certain metabolites—glucose, mannitol, fructose and pyruvate—are transported to the cytoplasm, some by their specific phosphotransferase system enzymes; they enter glycolysis, where their catabolism generates NADH, and this NADH is then oxidized by enzymes in the electron transport chain, which contribute to PMF. The elevated PMF facilitates the uptake of aminoglycosides, which bind to the ribosome and cause cell death via mistranslation.

We next investigated whether this mechanism was applicable to clinically relevant cases, such as bacterial biofilms. We reasoned that metabolic stimulation might facilitate the elimination of biofilm persisters by aminoglycosides. To test this hypothesis, we grew *E. coli* biofilms and treated them for 4 h with ofloxacin, mannitol, gentamicin and mannitol plus gentamicin (Fig. 3b). Ofloxacin, which is efficient against Gram-negative biofilms^{15,20}, reduced biofilm viability by almost two orders of magnitude, indicating that more than 1% of the bacteria in the biofilms were persisters. Mannitol and gentamicin, in combination, reduced biofilm viability by more than four orders of magnitude, representing a reduction in biofilm persisters of 2.5 orders of magnitude. We also tested the ability of fructose to induce biofilm elimination and observed similar results (Fig. 3c).

To determine the clinical relevance of metabolic potentiation of aminoglycosides *in vivo*, we tested the ability of gentamicin in combination with mannitol to treat chronic, biofilm-associated infection in a mouse model. Catheters colonized with uropathogenic *E. coli* biofilms were implanted in the urinary tracts of mice (Fig. 3d). Two days after surgery, mice received either no treatment or intravenous treatment with gentamicin or gentamicin and mannitol for 3 days, after which the catheters were removed and biofilm viability was determined. Gentamicin alone had no effect, whereas gentamicin in combination with mannitol reduced the viability of the catheter biofilms by nearly 1.5 orders of magnitude (Fig. 3d). We also found that treatment with gentamicin and mannitol inhibited the spread of bacterial infection to the kidneys,

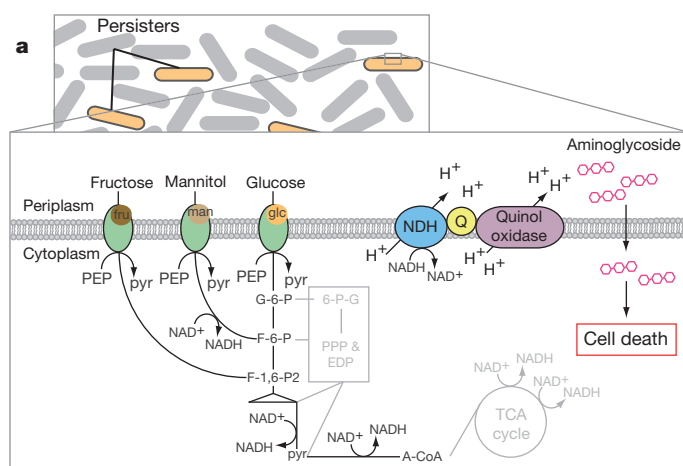
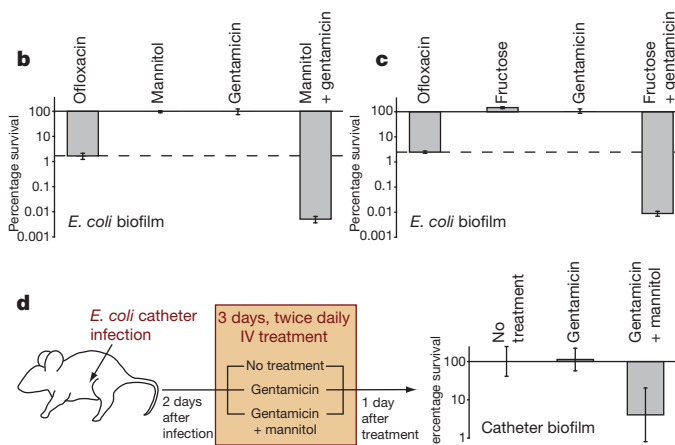


Figure 3 | Mechanism for metabolite-enabled eradication of persisters (a) and clinically relevant experiments (b–d). **a**, Metabolite-enabled eradication of persisters proceeds through catabolism of carbon sources, thereby generating NADH, the production of which does not require the PPP, EDP or TCA cycle. The electron transport chain oxidizes NADH and contributes to PMF, which facilitates aminoglycoside uptake and the killing of persisters. The quinones (Q), NADH dehydrogenases (NDH) and central metabolism, including pyruvate (pyr) and acetyl-CoA (A-CoA), are depicted. **b**, Percentage survival of *E. coli* biofilms after treatment with ofloxacin, mannitol, gentamicin or



mannitol plus gentamicin. Because quinolones have high efficacy against Gram-negative biofilms, as compared to other antibiotics^{15,20}, ofloxacin was used as a benchmark for high biofilm killing. **c**, Percentage survival of *E. coli* biofilms after treatment with ofloxacin, fructose, gentamicin or fructose plus gentamicin. **d**, Schematic of *in vivo* experiments in mice (left panel). The right panel shows survival of *E. coli* biofilms on urinary-tract-inserted catheters after treatment with gentamicin (1 mg kg^{-1}) or mannitol (1.5 g kg^{-1}) plus gentamicin (1 mg kg^{-1}). Means \pm s.e.m. are presented throughout ($n \geq 3$).

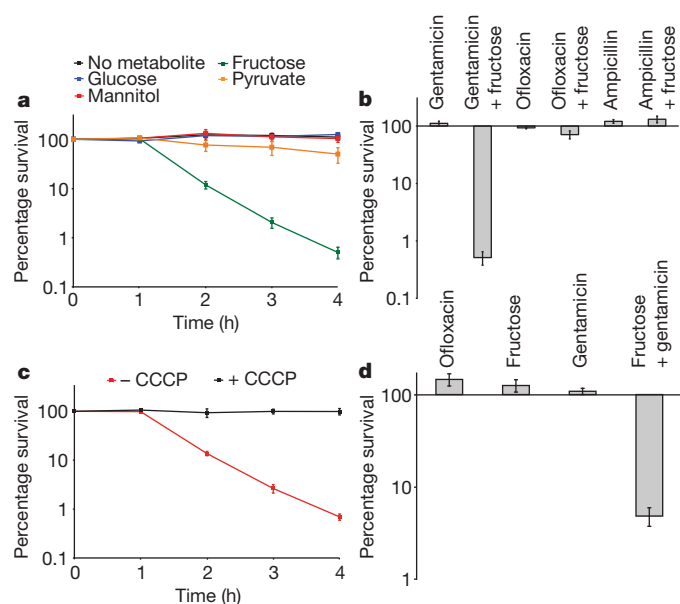


Figure 4 | Fructose induces PMF-dependent killing of *S. aureus* persisters by an aminoglycoside. **a**, Percentage survival of *S. aureus* persisters after treatment with gentamicin and different metabolites. **b**, Percentage survival of *S. aureus* persisters after 4-h treatment with fructose and different classes of antibiotics. **c**, Percentage survival of *S. aureus* persisters after 4-h treatment with gentamicin and fructose, with or without CCCP. **d**, Percentage survival of *S. aureus* biofilms after 4-h treatment with ofloxacin, fructose, gentamicin or fructose plus gentamicin. Means \pm s.e.m. are presented throughout ($n \geq 3$).

as compared to treatment with gentamicin alone and to the untreated control (Supplementary Fig. 27). These *in vivo* results demonstrate the feasibility of our approach for clinical use.

Having shown that certain metabolites can enable aminoglycoside activity in Gram-negative (*E. coli*) bacterial persisters and biofilms, we sought to determine whether a similar phenomenon existed in Gram-positive bacteria. Persisters of the Gram-positive pathogen *S. aureus* were treated with gentamicin in conjunction with metabolites. After an initial hour in which no killing was seen, gentamicin with fructose rapidly eliminated persistent *S. aureus* (Fig. 4a). Interestingly, mannitol, glucose and pyruvate, which showed strong potentiation against *E. coli* persisters, showed little potentiation against *S. aureus*. Expression analysis using *S. aureus* microarrays indicated that this lack of potentiation results from differential expression of metabolite transporters in growing versus dormant *S. aureus* (Supplementary Table 3). We next tested whether the fructose-enabled killing of *S. aureus* was unique to aminoglycosides or general to other classes of bactericidal antibiotics. As with *E. coli*, we found that metabolite-enabled killing of *S. aureus* persisters was specific to aminoglycosides (Fig. 4b), indicating that *S. aureus* persisters did not revert to normally growing cells when treated with metabolites.

Given that the activity of aminoglycosides in growing *S. aureus* is dependent on PMF^{21,22}, we tested whether the elimination of persisters mediated by fructose also required PMF. We found that the potentiation of aminoglycosides by fructose in *S. aureus*, as in *E. coli*, requires PMF generation (Fig. 4c), indicating that the PMF-requiring mechanism exists in both Gram-negative and Gram-positive bacteria. We also investigated whether gentamicin with fructose could be used to treat *S. aureus* biofilms. We found that the viability of *S. aureus* biofilms was reduced by nearly 1.5 orders of magnitude when treated for 4 h with fructose and gentamicin (Fig. 4d).

Here we have established an approach for eradicating persisters which is based on metabolism and is effective against both Gram-negative and Gram-positive bacteria. The metabolite-mediated potentiation occurs via PMF generation, which we found was necessary for aminoglycoside uptake and killing of persisters. This work adds

to a growing understanding of the role of metabolism in killing by bactericidal antibiotics^{13,23,24} and broadens our understanding of persister physiology. Moreover, our findings imply that delivering PMF-stimulating metabolites as adjuvants to aminoglycosides would be beneficial in the treatment of chronic bacterial infections.

METHODS SUMMARY

In all experiments, bacterial cells were cultured in 25 ml Luria-Bertani broth (LB) for 16 h at 37 °C, 300 r.p.m. and 80% humidity in 250 ml flasks. Unless otherwise noted, the following concentrations were used: 10 $\mu\text{g ml}^{-1}$ gentamicin, 100 $\mu\text{g ml}^{-1}$ ampicillin, 5 $\mu\text{g ml}^{-1}$ ofloxacin, 20 μM CCCP and 1 mM potassium cyanide. The concentration of all carbon sources added to potentiate aminoglycosides was normalized to deliver 60 mM carbon (for example, 10 mM glucose, 20 mM pyruvate, etc.). The parent strains used in this study were *E. coli* (K12 EMG2) and *S. aureus* (ATCC 25923). Knockouts (Supplementary Tables 1 and 2) were constructed by P1 phage transduction from the Keio knockout collection. In *E. coli*, non-persister cells in stationary phase were killed by treatment with 5 $\mu\text{g ml}^{-1}$ ofloxacin for 4 h (refs 25, 26). Samples were then washed with phosphate buffered saline (PBS) and suspended in M9 salts with a carbon source and antibiotic to determine metabolite-enabled killing of persisters. At specified time points, 10- μl aliquots were removed, serially diluted and spot-plated onto LB agar plates to determine colony-forming units per ml (c.f.u. ml^{-1}) and survival. Gentamicin-Texas red was made as previously described²⁷. Aminoglycoside uptake was measured by incubating stationary-phase samples with 10 $\mu\text{g ml}^{-1}$ gentamicin-Texas red for 5 min at 37 °C, 300 r.p.m. and 80% humidity. 100 μl of each sample was then washed, re-suspended in PBS and analysed on a BD FACS Aria II flow cytometer. Biofilm survival assays were performed as previously described²⁸. Raw microarray data for *S. aureus* were downloaded from the Gene Expression Omnibus series GSE20973 (ref. 29) and processed with RMA express using background adjustment, quantile normalization and median polish summarization to compute RMA expression values³⁰. Mouse experiments were performed with female Charles River BALB/c mice in collaboration with ViviSource Laboratories and conformed to the ViviSource institutional animal care and use policies and procedural guidelines.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 July 2010; accepted 24 March 2011.

- Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–1625 (2004).
- Gefen, O., Gabay, C., Mumcuoglu, M., Engel, G. & Balaban, N. Q. Single-cell protein induction dynamics reveals a period of vulnerability to antibiotics in persister bacteria. *Proc. Natl Acad. Sci. USA* **105**, 6145–6149 (2008).
- Gefen, O. & Balaban, N. Q. The importance of being persistent: heterogeneity of bacterial populations under antibiotic stress. *FEMS Microbiol. Rev.* **33**, 704–717 (2009).
- Lewis, K. Persister cells, dormancy and infectious disease. *Nature Rev. Microbiol.* **5**, 48–56 (2007).
- Smith, P. A. & Romesberg, F. E. Combating bacteria and drug resistance by inhibiting mechanisms of persistence and adaptation. *Nature Chem. Biol.* **3**, 549–556 (2007).
- Levin, B. R. & Rozen, D. E. Non-inherited antibiotic resistance. *Nature Rev. Microbiol.* **4**, 556–562 (2006).
- Dhar, N. & McKinney, J. D. Microbial phenotypic heterogeneity and antibiotic tolerance. *Curr. Opin. Microbiol.* **10**, 30–38 (2007).
- Shah, D. *et al.* Persisters: a distinct physiological state of *E. coli*. *BMC Microbiol.* **6**, 53 (2006).
- Vakulenko, S. B. & Mobashery, S. Versatility of aminoglycosides and prospects for their future. *Clin. Microbiol. Rev.* **16**, 430–450 (2003).
- Magnet, S. & Blanchard, J. S. Molecular insights into aminoglycoside action and resistance. *Chem. Rev.* **105**, 477–498 (2005).
- Davis, B. D. Mechanism of bactericidal action of aminoglycosides. *Microbiol. Rev.* **51**, 341–350 (1987).
- Weisblum, B. & Davies, J. Antibiotic inhibitors of the bacterial ribosome. *Bacteriol. Rev.* **32**, 493–528 (1968).
- Kohanski, M. A., Dwyer, D. J., Wierzbowski, J., Cottarel, G. & Collins, J. J. Mistranslation of membrane proteins and two-component system activation trigger antibiotic-mediated cell death. *Cell* **135**, 679–690 (2008).
- Keren, I., Shah, D., Spoering, A., Kaldalu, N. & Lewis, K. Specialized persister cells and the mechanism of multidrug tolerance in *Escherichia coli*. *J. Bacteriol.* **186**, 8172–8180 (2004).
- Spoering, A. L. & Lewis, K. Biofilms and planktonic cells of *Pseudomonas aeruginosa* have similar resistance to killing by antimicrobials. *J. Bacteriol.* **183**, 6746–6751 (2001).
- Taber, H. W., Mueller, J. P., Miller, P. F. & Arrow, A. S. Bacterial uptake of aminoglycoside antibiotics. *Microbiol. Rev.* **51**, 439–457 (1987).

17. Bryan, L. E. & Kwan, S. Roles of ribosomal binding, membrane potential, and electron transport in bacterial uptake of streptomycin and gentamicin. *Antimicrob. Agents Chemother.* **23**, 835–845 (1983).
18. Hill, S., Viollet, S., Smith, A. T. & Anthony, C. Roles for enteric d-type cytochrome oxidase in N₂ fixation and microaerobiosis. *J. Bacteriol.* **172**, 2071–2078 (1990).
19. Govantes, F., Albrecht, J. A. & Gunsalus, R. P. Oxygen regulation of the *Escherichia coli* cytochrome d oxidase (cydAB) operon: roles of multiple promoters and the Fnr-1 and Fnr-2 binding sites. *Mol. Microbiol.* **37**, 1456–1469 (2000).
20. Walters, M. C. III, Roe, F., Bugnicourt, A., Franklin, M. J. & Stewart, P. S. Contributions of antibiotic penetration, oxygen limitation, and low metabolic activity to tolerance of *Pseudomonas aeruginosa* biofilms to ciprofloxacin and tobramycin. *Antimicrob. Agents Chemother.* **47**, 317–323 (2003).
21. Mates, S. M. *et al.* Membrane potential and gentamicin uptake in *Staphylococcus aureus*. *Proc. Natl Acad. Sci. USA* **79**, 6693–6697 (1982).
22. Fraimow, H. S., Greenman, J. B., Leviton, I. M., Dougherty, T. J. & Miller, M. H. Tobramycin uptake in *Escherichia coli* is driven by either electrical potential or ATP. *J. Bacteriol.* **173**, 2800–2808 (1991).
23. Dwyer, D. J., Kohanski, M. A., Hayete, B. & Collins, J. J. Gyrase inhibitors induce an oxidative damage cellular death pathway in *Escherichia coli*. *Mol. Syst. Biol.* **3**, 91 (2007).
24. Kohanski, M. A., Dwyer, D. J., Hayete, B., Lawrence, C. A. & Collins, J. J. A common mechanism of cellular death induced by bactericidal antibiotics. *Cell* **130**, 797–810 (2007).
25. Keren, I., Kaldalu, N., Spoering, A., Wang, Y. & Lewis, K. Persister cells and tolerance to antimicrobials. *FEMS Microbiol. Lett.* **230**, 13–18 (2004).
26. Hansen, S., Lewis, K. & Vulic, M. Role of global regulators and nucleotide metabolism in antibiotic tolerance in *Escherichia coli*. *Antimicrob. Agents Chemother.* **52**, 2718–2726 (2008).
27. Sandoval, R., Leiser, J. & Molitoris, B. A. Aminoglycoside antibiotics traffic to the Golgi complex in LLC-PK1 cells. *J. Am. Soc. Nephrol.* **9**, 167–174 (1998).
28. Lu, T. K. & Collins, J. J. Dispersing biofilms with engineered enzymatic bacteriophage. *Proc. Natl Acad. Sci. USA* **104**, 11197–11202 (2007).
29. Majerczyk, C. D. *et al.* Direct targets of CodY in *Staphylococcus aureus*. *J. Bacteriol.* **192**, 2861–2877 (2010).
30. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Slee and T. Murphy from ViviSource Laboratories for assistance with the *in vivo* mouse studies and T. K. Lu for guidance with the biofilm experiments. This work was supported by the NIH Director's Pioneer Award Program and the Howard Hughes Medical Institute.

Author Contributions All authors designed the study, analysed results and wrote the manuscript. Experiments were performed by K.R.A. and M.P.B.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.J.C. (jjcollins@bu.edu).

METHODS

Antibiotics and chemicals. The following concentrations of antibiotics were used in this study: 10 µg ml⁻¹ gentamicin, 30 µg ml⁻¹ kanamycin, 50 µg ml⁻¹ streptomycin, 5 µg ml⁻¹ ofloxacin, 100 µg ml⁻¹ ampicillin, 40 µg ml⁻¹ tetracycline, 50 µg ml⁻¹ chloramphenicol and 100 µg ml⁻¹ spectinomycin. 20 µM carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP) was used in experiments to suppress proton-motive force. 1 mM potassium cyanide was used to inhibit aerobic respiration. A stock solution of CCCP was made in dimethylsulphoxide at 500 µM and stored at 4 °C. All antibiotics and chemicals were purchased from Sigma and Fisher.

Media and growth conditions. All stationary-phase samples were prepared in the following way: cells from frozen stock were grown at 37 °C, 300 r.p.m. and 80% humidity in Luria-Bertani (LB) broth to an optical density at 600 nm (OD₆₀₀) of 0.3. Cells were then diluted 1:1,000 in 25 ml LB and grown for 16 h at 37 °C, 300 r.p.m. and 80% humidity in 250 ml flasks. This experimental set-up ensured that, regardless of the conditions used in assays, the initial population of persisters or stationary-phase cells was prepared in a uniform manner consistent with previous work^{25,26}.

The following concentrations of carbon sources were used in this study: 10 mM glucose, 10 mM mannitol, 10 mM fructose, 20 mM glycerol, 30 mM glycolate, 10 mM galactarate, 20 mM pyruvate, 10 mM gluconate, 12 mM arabinose and 12 mM ribose. 10 mM acetate was used as a supplement and 30 mM acetate was used as a potentiating metabolite in control experiments. Carbon sources were purchased from Sigma or Fisher. In *E. coli* persister assays, samples were re-suspended in M9 minimal media salts plus a carbon source. In all experiments using CCCP, samples were pre-treated for 5 min with the proton ionophore before antibiotic addition.

Strains. The parent strains used in this study were *E. coli* (K12 EMG2) and *S. aureus* (ATCC 25923). All *E. coli* knockouts (see Supplementary Table 2) were transduced into the EMG2 strain from strains in the KEIO knockout collection using the P1 phage method. All strains were cured using pCP20 and standard laboratory procedures before their use in assays.

Persister assays. For *E. coli* persister assays, samples were grown to stationary phase as described above. Cultures were then treated for 4 h with 5 µg ml⁻¹ ofloxacin in the above-stated growth conditions. Previous work has demonstrated that treatment for 3 h under these conditions eliminates all susceptible non-persister cells²⁵. We verified that the remaining cells were persisters by increasing the amount of ofloxacin added to the culture to 20 µg ml⁻¹ and noted no further decrease in viability (see Supplementary Fig. 1). Persistence was further verified by demonstrating that none of the bactericidal antibiotics used in this study killed cells when added without a carbon source (see Fig. 1c). Samples were then washed with 10 ml of 1× filtered PBS and re-suspended in M9 minimal media. Carbon sources and antibiotics were added and samples were incubated at 37 °C, 300 r.p.m. and 80% humidity. Re-suspension of samples in defined minimal media allowed us to test the effects of different carbon sources on persister viability precisely and specifically, without the possible confounding factors that a more complex medium would present. For *S. aureus* persister experiments, cells were grown at 37 °C and 300 r.p.m. in LB broth to an OD₆₀₀ of 0.3. Cells were then diluted 1:1,000 in 25 ml LB and grown for 16 h at 37 °C and 300 r.p.m. in 250 ml flasks before treatment with a carbon source and antibiotic. Previous work has shown that almost all stationary-phase *S. aureus* cells are persistent²⁵.

At specified time points, 10-µl aliquots of samples were removed, serially diluted and spot-plated onto LB agar plates to determine colony-forming units per ml (c.f.u. ml⁻¹). Only dilutions that yielded 10–100 colonies were counted. Survival was determined by dividing the c.f.u. ml⁻¹ of a sample at each time point by the initial c.f.u. ml⁻¹ for that sample.

For persister resuscitation experiments (Supplementary Fig. 5), cells that had been treated with ofloxacin, washed and re-suspended in M9 were then diluted 1:100 in M9 plus a carbon source and incubated at 37 °C, 300 r.p.m. and 80% humidity.

Gentamicin-Texas red uptake. Gentamicin-Texas red was made as previously described²⁷. At 4 °C, 1 mg of Texas red (Invitrogen) was dissolved in 50 µl of high-quality anhydrous *N,N*-dimethylformamide. The dissolved Texas red was added slowly to 2.3 ml of 10 mg ml⁻¹ gentamicin solution in 100 mM K₂CO₃ at 4 °C.

Uptake of gentamicin-Texas red induced by carbon sources in stationary-phase cells was determined by adding concentrated carbon sources (see section on media and growth conditions for concentrations) to stationary-phase cultures and incubating them for 30 min at 37 °C, 300 r.p.m. and 80% humidity. Concentrated

gentamicin-Texas red was then added to a final concentration of 10 µg ml⁻¹ and samples were incubated for 5 min. 100 µl of each sample was then washed with 1 ml of PBS and re-suspended in 1 ml of PBS. 200 µl of the re-suspended sample was then added to 800 µl of PBS in flow tubes. Samples were analysed on a BD FACS Aria II flow cytometer with the following settings: mCherry voltage, 650 V; FSC threshold, 1,000; recorded events, 100,000; gated-out mCherry events, <30.

Membrane potential measurements. We used the BacLight Bacterial Membrane Potential Kit (B34950, Invitrogen) to assess changes in proton-motive force induced by metabolites. Membrane potential induced by carbon sources in stationary-phase cells was determined by adding concentrated carbon sources (see section on media and growth conditions for concentrations) and 10 µl DiOC₂(3) (membrane stain) to stationary-phase cultures, followed by incubation for 30 min at 37 °C, 300 r.p.m. and 80% humidity. 10 µl of culture was added to 1 ml of PBS in flow tubes immediately before analysing. Samples were analysed on a BD FACS Aria II flow cytometer with settings optimized according to the BacLight kit manual. Settings used were: FITC voltage, 250 V; mCherry voltage, 650 V; FSC threshold, 1,000; recorded events, 100,000. FSC and SSC outliers were gated out by visual inspection before data acquisition. The red/green (mCherry/FITC) values for each cell were determined and normalized, then compared to samples without metabolite to determine the relative change in PMF.

Anaerobic experiments. *Escherichia coli* were grown to an OD₆₀₀ of 0.3, then diluted 1:1,000 in 25 ml anaerobic LB with 10 mM NaNO₃ and grown for 16 h at 37 °C, 200 r.p.m., 1.5–2.0% hydrogen and <50 p.p.m. oxygen in 250 ml flasks. Cultures were then treated with a carbon source and metabolite in the presence or absence of an additional 10 mM NaNO₃. The additional NaNO₃ was added to determine whether increasing the concentration of the terminal electron acceptor could increase aminoglycoside potentiation.

Biofilm assay. Overnight cultures grown in LB were diluted 1:200 into pre-warmed LB which was then added to MBEC plates (Innovotech) at 150 µl per well. Plates were incubated at 35 °C, 150 r.p.m. for 24 h, then pegs were washed in a microtitre plate with 200 µl of 1× PBS per well. Pegs were then added to a microtitre plate containing 200 µl M9 minimal salts (for *E. coli*) or sterile-filtered, stationary-phase media (for *S. aureus*), plus a carbon source and antibiotic. Plates were incubated at 35 °C, 150 r.p.m. for 4 h, then pegs were washed twice in microtitre plates with 200 µl of 1× PBS per well. To dislodge biofilm cells, pegs were placed in a microtitre plate with 145 µl of 1× PBS per well and sonicated in a water bath for 30 min at 40 kHz. Serial dilutions and spot-plating were performed to determine viable c.f.u. per peg. For determination of the dependence of *E. coli* biofilm elimination on pH, we carried out the above procedure in M9 salts buffered to an appropriate pH with citric acid, as opposed to KH₂PO₄, which is typically used for M9.

Mouse chronic urinary tract infection assay. Female Charles River BALB/c mice (weighing 22–26 g) received surgical implantation in the urinary tract of 6 mm PE50 catheter tubing that had been incubated in cultures of uropathogenic *E. coli* for 24 h to form biofilms. 48 h after surgery, mice received either no treatment or twice-daily, intravenous treatment with gentamicin (1 mg kg⁻¹) or mannitol (1.5 g kg⁻¹) plus gentamicin (1 mg kg⁻¹) for 3 days. Seven or eight mice were included in each group. 24 h after the last treatment, catheter tubing was extracted to determine biofilm viability and kidneys were removed to determine bacterial load. Mouse materials were provided by ViviSource Laboratories, a facility approved by the US Department of Agriculture and by the Office of Laboratory Animal Welfare, where all *in vivo* experimental work was performed. The study conformed with ViviSource institutional animal care and use policies and procedural guidelines.

Staphylococcus aureus microarray analysis. Raw microarray data (.CEL files) for two exponential (GSM524189, GSM524193) and two stationary phase (GSM524362, GSM524363) *S. aureus* cultures were downloaded from the Gene Expression Omnibus (GEO) series GSE20973 (ref. 29). The data were processed with RMA express using background adjustment, quantile normalization and median polish summarization to compute RMA expression values³⁰. Mean expression values were calculated for both exponential and stationary-phase data and the relative fold changes (stationary/exponential) are reported in Supplementary Table 3.

Software. MATLAB (Mathworks) was used for processing flow cytometric data, analysing microarray data and generating scaled heat maps using the imagesc function. Microsoft Excel was used to plot survival assays. All figures were formatted with Adobe Illustrator.

Modelling schizophrenia using human induced pluripotent stem cells

Kristen J. Brennand¹, Anthony Simone^{1*}, Jessica Jou^{1*}, Chelsea Gelboin-Burkhart^{1*}, Ngoc Tran^{1*}, Sarah Sangar¹, Yan Li¹, Yangling Mu¹, Gong Chen², Diana Yu¹, Shane McCarthy³, Jonathan Sebat⁴ & Fred H. Gage¹

Schizophrenia (SCZD) is a debilitating neurological disorder with a world-wide prevalence of 1%; there is a strong genetic component, with an estimated heritability of 80–85%¹. Although post-mortem studies have revealed reduced brain volume, cell size, spine density and abnormal neural distribution in the prefrontal cortex and hippocampus of SCZD brain tissue² and neuropharmacological studies have implicated dopaminergic, glutamatergic and GABAergic activity in SCZD³, the cell types affected in SCZD and the molecular mechanisms underlying the disease state remain unclear. To elucidate the cellular and molecular defects of SCZD, we directly reprogrammed fibroblasts from SCZD patients into human induced pluripotent stem cells (hiPSCs) and subsequently differentiated these disorder-specific hiPSCs into neurons (Supplementary Fig. 1). SCZD hiPSC neurons showed diminished neuronal connectivity in conjunction with decreased neurite number, PSD95-protein levels and glutamate receptor expression. Gene expression profiles of SCZD hiPSC neurons identified altered expression of many components of the cyclic AMP and WNT signalling pathways. Key cellular and molecular elements of the SCZD phenotype were ameliorated following treatment of SCZD hiPSC neurons with the antipsychotic loxapine. To date, hiPSC neuronal pathology has only been demonstrated in diseases characterized by both the loss of function of a single gene product and rapid disease progression in early childhood^{4–6}. We now report hiPSC neuronal phenotypes and gene expression changes associated with SCZD, a complex genetic psychiatric disorder.

Four SCZD patients were selected: patient 1, diagnosed at 6 years of age, had childhood-onset SCZD; patients 2, 3 and 4 were from families in which all offspring and one parent were affected with psychiatric disease (Supplementary Fig. 3a). Primary human fibroblasts were reprogrammed using inducible lentiviruses⁷. Control and SCZD hiPSCs expressed endogenous pluripotency genes, repressed viral genes and were indistinguishable in assays for self-renewal and pluripotency (Fig. 1 and Supplementary Fig. 2). SCZD hiPSCs had no apparent defects in generating neural progenitor cells (NPCs) or neurons (Fig. 1 and Supplementary Fig. 3). Most hiPSC neurons were presumably glutamatergic and expressed VGLUT1 (also known as SLC17A7; Supplementary Fig. 8a). Approximately 30% of neurons were GAD65/67-positive (also known as GAD1/2; GABAergic) (Supplementary Fig. 8c, d) whereas less than 10% of neurons were tyrosine hydroxylase-positive (dopaminergic) (Supplementary Fig. 7).

Neuronal connectivity was assayed using trans-neuronal spread of rabies; *in vivo*, rabies transmission occurs via synaptic contacts and is strongly correlated with synaptic input strength⁸. Primary infection was restricted by replacing the rabies coat protein with envelope A (ENVA), which infects only via the avian tumour virus A (TVA) receptor; viral spread was limited to monosynaptically connected neurons by deleting the rabies glycoprotein gene (ΔG)⁹. Neurons were first transduced with a lentivirus expressing histone 2B (H2B)–green fluorescent protein

(GFP) fusion protein, TVA and G from the synapsin (SYN) promoter (LV-SYNP-HTG). One week later, neurons were transduced with modified rabies virus (Rabies-ENVA ΔG -RFP). Primary infected cells were positive for both H2B–GFP and RFP (red fluorescent protein); neurons monosynaptically connected to primary cells were GFP-negative but RFP-positive (Supplementary Fig. 4a). Transduction with Rabies-ENVA ΔG -RFP alone resulted in no RFP-positive cells, whereas transduction with Rabies-ENVA ΔG -RFP following lentiviral transduction without rabies glycoprotein (SYNP-HT) led to only single GFP⁺RFP⁺ cells, indicating that *in vitro* rabies infection and spread are dependent on TVA expression and G trans-complementation, respectively (Supplementary Fig. 4c, d).

There was decreased neuronal connectivity in SCZD hiPSC neurons (Fig. 2; Supplementary Figs 4b, c, 5 and 6). Fluorescence-activated cell sorting (FACS) analysis confirmed differences in neuronal connectivity and demonstrated that comparable numbers of β III-tubulin-positive neurons were labelled with LV-SYNP-HTG. Though the mechanism of rabies trans-neuronal tracing is not fully understood, the presynaptic protein NCAM has been implicated¹⁰; NCAM expression is decreased in SCZD hiPSC neurons (Supplementary Table 3). Rabies trans-neuronal tracing occurs in functionally immature hiPSC neurons (Supplementary Fig. 4e) and in the presence of the voltage-gated sodium channel blocker tetrodotoxin (TTX) (1 μ M), depolarizing KCl (50 mM) or the calcium channel blocker ryanodine (10 μ M) (Supplementary Fig. 4f). Decreased

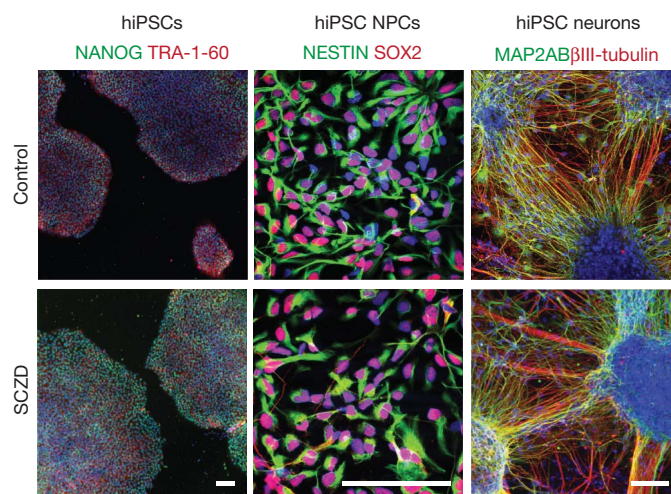


Figure 1 | Patient-specific hiPSCs, NPCs and neurons. Left, hiPSCs express NANOG (green) and TRA-1-60 (red). DAPI (blue). $\times 100$, scale bar 2.5 μ m. Centre, hiPSC neural progenitor cells (NPCs) express NESTIN (green) and SOX2 (red). DAPI (blue). $\times 630$, scale bar 2.5 μ m. Right, hiPSC neurons express β III-tubulin (red) and the dendritic marker MAP2AB (green). DAPI (blue). $\times 200$, scale bar 2.5 μ m.

¹Salk Institute for Biological Studies, Laboratory of Genetics, 10010 North Torrey Pines Road, La Jolla California 92037, USA. ²Department of Biology, Pennsylvania State University, 201 Life Science Building, University Park, Pennsylvania 16802, USA. ³Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. ⁴University of California San Diego, Department of Psychiatry and Department of Cellular and Molecular Medicine, University Of California, San Diego, La Jolla, California 92093, USA.

*These authors contributed equally to this work.

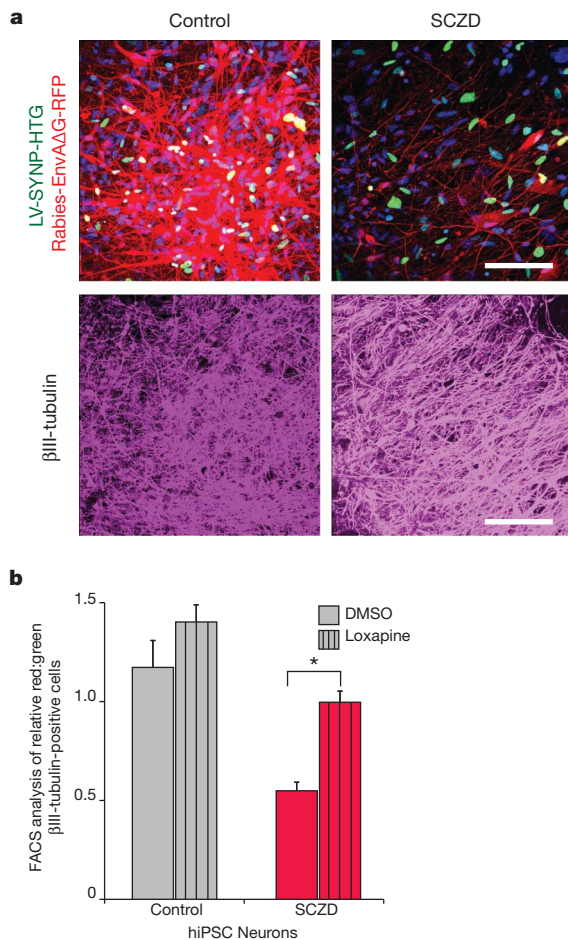


Figure 2 | Decreased neuronal connectivity in SCZD hiPSC neurons.

a, Representative images of control and SCZD hiPSC neurons cotransduced with LV-SYNP-HTG and Rabies-EnvAΔG-RFP, 10 days post rabies transduction. All images were captured using identical laser power and gain settings. βIII-tubulin staining (purple) of the field is shown below each panel. $\times 400$, scale bar 2 μm . **b**, Graph showing treatment of SCZD hiPSC neurons with Loxapine resulted in a statistically significant improvement in neuronal connectivity. Error bars are s.e., $*P < 0.05$

trans-neuronal tracing is evidence of decreased neuronal connectivity, but not necessarily decreased synaptic function, in SCZD hiPSC neurons.

We tested the ability of five antipsychotic drugs to improve neuronal connectivity *in vitro*. Clozapine, loxapine, olanzapine, risperidone and thioridazine were administered for the final 3 weeks of neuronal differentiation. Only loxapine significantly increased neuronal connectivity in hiPSC neurons from all patients (Fig. 2b and Supplementary Fig. 5). Optimization of the concentration and timing of drug administration may improve the effects of the other antipsychotic medications.

Reduced dendritic arborization has been observed in post-mortem SCZD brains¹¹ and in animal models¹². SCZD hiPSC neurons show a decrease in the number of neurites (Fig. 3a and Supplementary Fig. 9a, b). Synaptic genes are associated with SCZD¹³ (Supplementary Fig. 9d) and impaired synaptic maturation occurs in a number of mouse models¹². hiPSC neurons express dense puncta of synaptic markers that co-stain for both pre- and post-synaptic markers (Supplementary Fig. 8a, b). Whereas we observed decreased PSD95 protein (also known as DLG4, levels relative to microtubule associated protein 2ab (MAP2AB) in SCZD hiPSC neuronal dendrites (Fig. 3b; Supplementary Fig. 9h), the levels of SYN, VGLUT1, GLUR1 (also known as GRIA1), VGAT (also known as SLC32A1) and GEPH (also known as GPHN) were unaffected (Supplementary Fig. 9e–j). Decreased PSD95 synaptic density

in SCZD hiPSC neurons failed to reach statistical significance (Fig. 3c and Supplementary Fig. 9c).

We used electrophysiology and calcium transient imaging to measure spontaneous neuronal activity (Fig. 3d–k and Supplementary Fig. 10). SCZD hiPSC neurons showed normal transient inward sodium currents and sustained outward potassium currents in response to membrane depolarizations (Fig. 3d), action potentials to somatic current injections (Fig. 3e), spontaneous excitatory postsynaptic currents (EPSCs) (Fig. 3f) and spontaneous inhibitory postsynaptic currents (IPSCs) (Fig. 3g). The amplitude and rate of spontaneous calcium transients were unaffected (Fig. 3h–j and Supplementary Fig. 10a–d) and there was no difference in synchronicity of spontaneous calcium transients (Fig. 3k and Supplementary Fig. 10e–g).

Increased *NRG1* expression has been observed in post-mortem SCZD brain tissue¹³. *NRG1* expression was increased in SCZD hiPSC neurons (Fig. 4d–f) but not SCZD fibroblasts, hiPSCs or NPCs (Fig. 4e), demonstrating the importance of studying gene expression changes in the cell type relevant to disease. In all, 596 unique genes (271 upregulated and 325 downregulated) showed greater than 1.30-fold-expression changes between SCZD and control hiPSC neurons ($P < 0.05$) (Supplementary Fig. 11a, b and Supplementary Table 3). Of these genes, 13% (74) have published associations with SCZD and 16% (96) have been linked to SCZD by post-mortem gene expression profiles available through the Stanley Medical Research Institute¹⁴ (Supplementary Table 3); in total 25% (149) of our differentially expressed genes have been previously implicated in SCZD. Gene ontology (GO) analysis identified significant perturbations of glutamate, cAMP and WNT signalling (Fig. 4a–c, Supplementary Table 4 and Supplementary Fig. 11c), pathways required for activity-dependent refinement of synaptic connections and long-term potentiation^{15–17}. Sixteen of 17 candidate genes from these families were validated by quantitative PCR (Supplementary Table 2; Fig. 4f and Supplementary Fig. 11e).

Copy number variants (CNVs) are rare, highly penetrant structural disruptions. SCZD patients have a 1.15-fold increase in CNV burden, but how this translates into illness is unknown. Patient 4 had four CNVs involving genes previously associated with SCZD or bipolar disorder (BD)^{13,18,19}; of these, neuronal expression of *NRG3* and *GALNT11*, but not of *CYP2C19* or *GABRB2/GABRA6* (also known as *GABRB2* and *GABRA6*, respectively) was affected (Supplementary Fig. 12 and Supplementary Table 5). A second analysis of CNVs unbiased by previous genome wide association studies (GWAS) identified 42 genes affected by CNVs in our four SCZD patients (Supplementary Table 5). Although twelve of these genes showed altered neuronal expression consistent with genotype ($P < 0.05$), most changes were extremely small and only three (*CSMD1*, *MYH1*, *MYH4*) showed >1.3 -fold effects (Supplementary Table 5). Well-established SCZD CNVs occur at 1q21.1, 15q11.2, 15q13.3, 16p11.2 and 22q11.2 (refs 13, 18, 19), but the relevant genes remain unidentified. Our patients had no evidence of CNVs at these regions, and gene expression of the best candidate genes in each region, such as *GJA8* (1q21.1), *CYFIP1* (15q11.1), *CHRFAM7A* (15q13.3), *PRODH* (22q11.2), *COMT* (22q11.2) and *ZDHHC8* (22q11.2)^{18,20}, was not affected in our SCZD hiPSC neurons (Supplementary Table 6).

Consistent with published reports, loxapine increased *NRG1* expression in neurons²¹. Loxapine also increased expression of several glutamate receptors, including *GRIK1*, *GRM7* and *GRIN2A*, and ameliorated expression of *ADCY8*, *PRKCA*, *WNT7A* and *TCF4* (Fig. 4f and Supplementary Fig. 11e).

SCZD hiPSC neurons from heterogeneous patients had similar deficits, replicating some but not all aspects of the cellular and molecular phenotypes observed in post-mortem human studies and animal models (Supplementary Table 1). We observed decreased neuronal connectivity in SCZD hiPSC neurons, but not defects in synaptic function; this may reflect technical limitations of our synaptic activity assays. Due to the heterogeneity of our patient cohort and small sample size, our findings might not generalize to all subtypes of SCZD and our microarray

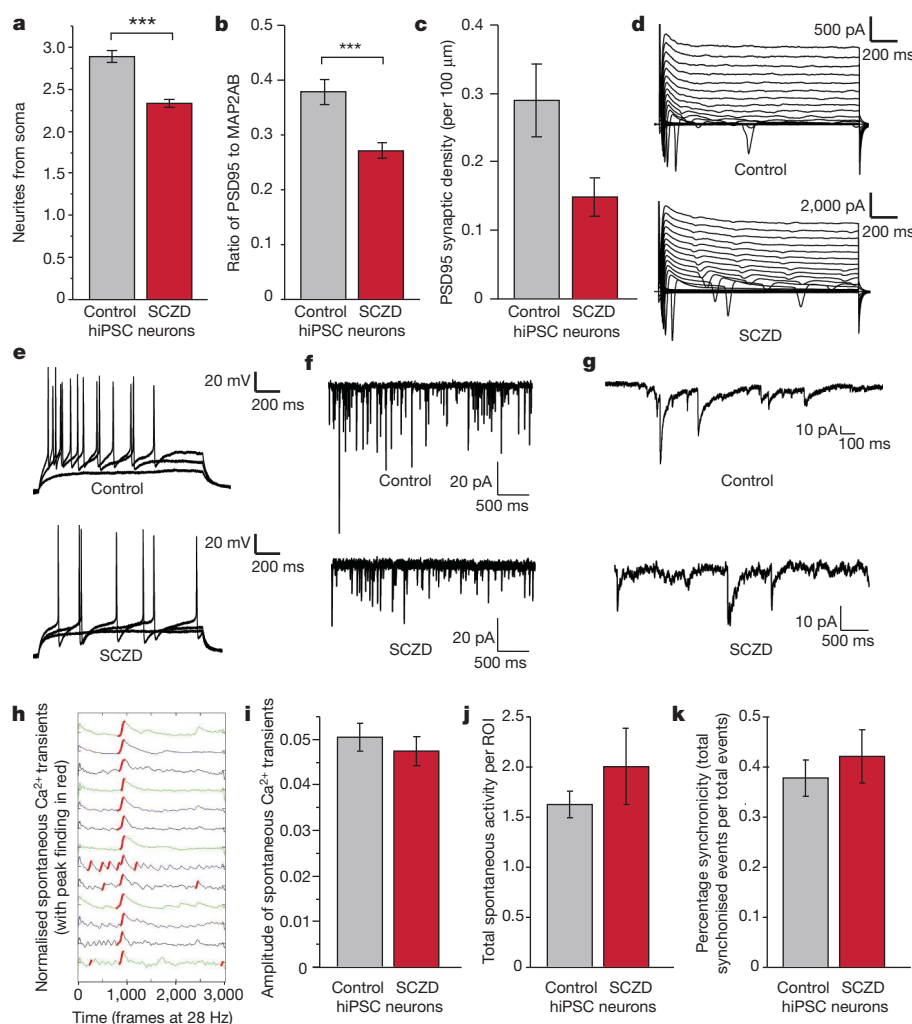


Figure 3 | Decreased neurites and synaptic protein levels but normal electrophysiological and spontaneous calcium transient activity in SCZD hiPSC neurons. **a**, Decreased neurites in SCZD hiPSC neurons. **b**, Decreased PSD95 protein relative to MAP2AB for SCZD hiPSC neurons. **c**, Trend of decreasing PSD95 synaptic density in SCZD hiPSC neurons. **d–g**, Electrophysiological characterization. hiPSC neurons grown on astrocytes show normal sodium and potassium currents when voltage-clamped (**d**), normal induced action potentials when current-clamped (**e**), and spontaneous excitatory (**f**) and inhibitory (**g**) synaptic activity.

h–k, Spontaneous calcium transient imaging. Representative spontaneous Fluo-4AM calcium traces of fluorescent intensity versus time generated from 3-month-old hiPSC neurons (**h**). There is no difference between the spike amplitude of spontaneous calcium transients of control and SCZD hiPSC neurons (**i**), no difference between the total numbers of spontaneous calcium transients per total number of regions of interest in cultures of control and SCZD hiPSC neurons (**j**), and no change in percentage synchronicity per calcium transient in control and SCZD hiPSC neurons (**k**). Error bars are s.e., *** $P < 0.001$.

comparisons of SCZD and control hiPSC neurons are necessarily preliminary. Gene expression studies of hiPSC neurons permit straightforward comparisons of antipsychotic treatments on live, genetically identical neurons from patients with known clinical treatment outcomes, eliminating many confounding variables of post-mortem analysis such as treatment history, drug or alcohol abuse, and cause of death. For example, although loxapine is characterized as a high affinity antagonist of serotonin 5-HT₂ receptors and dopamine D1, D2 and D4 receptors²², treatment of SCZD hiPSC neurons resulted in altered gene expression and increased neuronal connectivity.

Of the 596 unique genes differentially expressed in our SCZD hiPSC neurons (>1.30-fold, $P < 0.05$), 25% have been previously implicated in SCZD (Supplementary Table 3). Although our gene expression profiles of SCZD hiPSC neurons confirm and extend the major hypotheses generated by pharmacological and GWAS studies of SCZD, they also identify some pathways not before linked to SCZD, such as NOTCH signalling, SLIT/ROBO axon guidance, EFNA mediated

axon growth, cell adhesion and transcriptional silencing (Supplementary Table 4). Many of the genes most affected in SCZD hiPSC neurons belong to pathways previously associated with SCZD, although they have not yet been singled out as SCZD genes. For example, whereas *PDE4B* is a well-characterized SCZD gene, we observed significant misexpression of *PDE1C*, *PDE3A*, *PDE4D*, *PDE4DIP*, *PDE7B*, *ADCY7* and *ADCY8*. Additionally, although some key SCZD/BD genes, including *NRG1* and *ANK3*, were misexpressed in all of our SCZD hiPSC neurons, many others, including *ZNF804A*, *GABRB1*, *ERBB4*, *DISC1* and *PDE4B*, were aberrantly expressed in some but not all patients (Fig. 4d). Our data support the “watershed model”²³ of SCZD whereby many different combinations of gene misfunction may disrupt the key pathways affected in SCZD. We predict that, as the number of SCZD cases studied using hiPSC neurons increases, a diminishing number of genes will be consistently affected across the growing patient cohort; instead, evidence will accumulate that a handful of essential pathways can be disrupted in diverse ways to result in SCZD.

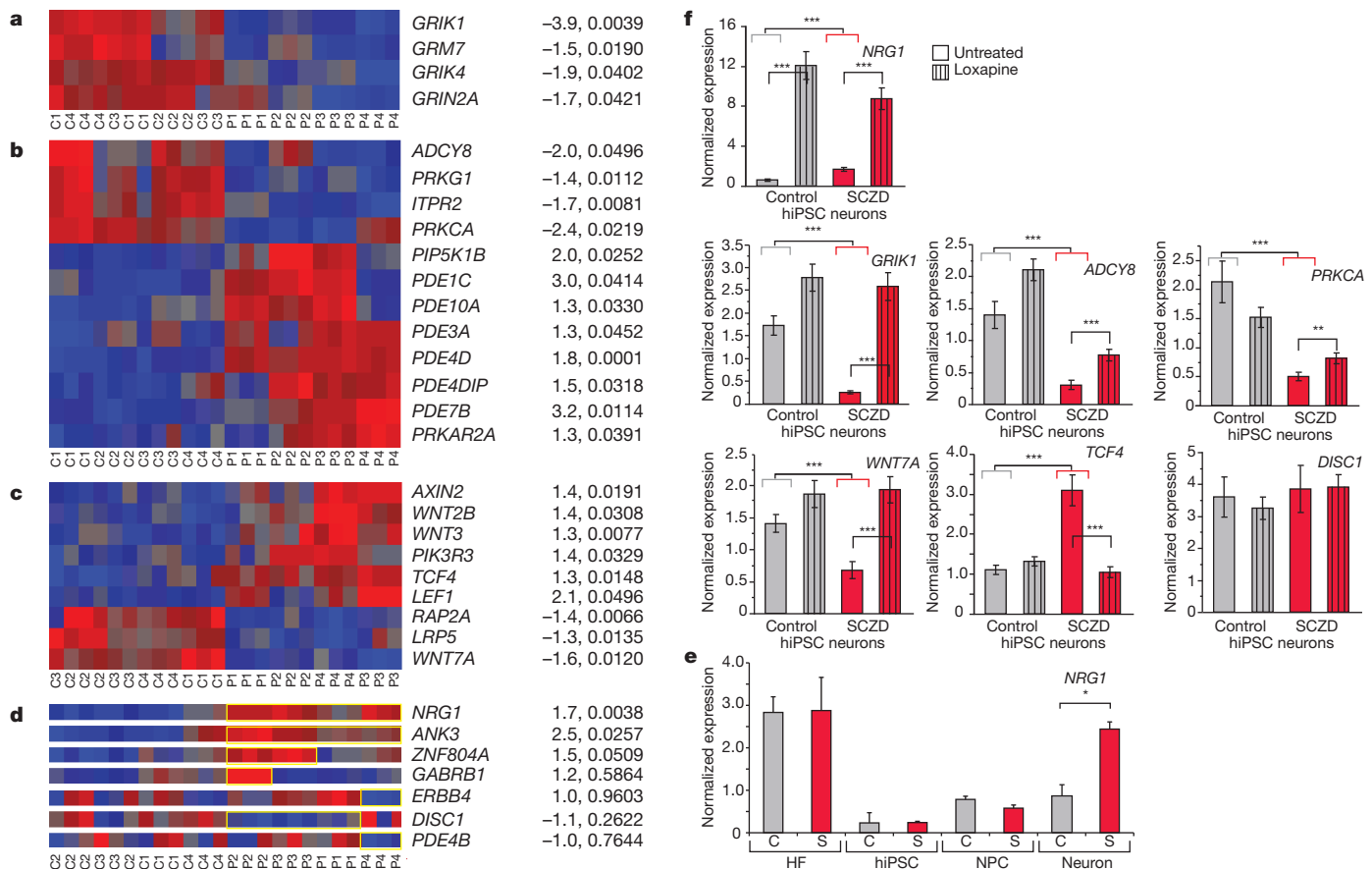


Figure 4 | RNA expression analysis of control and SCZD hiPSC neurons. **a–c**, Heat maps showing microarray expression profiles of altered expression of glutamate receptors (**a**), cAMP signalling (**b**) and WNT signalling (**c**) genes in SCZD hiPSC neurons. Fold-change and *P*-values (diagnosis) are shown on the right. **d**, Heat maps showing perturbed expression (highlighted in yellow) of *NRG1* and *ANK3* in all four SCZD patients, as well as altered expression of *ZNF804A*, *GABRB1*, *ERBB4*, *DISC1* and *PDE4B* in some but not all patients.

METHODS SUMMARY

Reprogramming hiPSCs. Control and SCZD human fibroblasts were obtained from cell repositories and were reprogrammed with tetracycline-inducible lentiviruses expressing the transcription factor genes *OCT4* (also known as *POU5F1*), *SOX2*, *KLF4*, *cMYC* and *LIN28* (ref. 7). Lentiviruses were packaged in 293T HEK cells transfected with polyethylenimine (PEI) (Polysciences). Human fibroblasts were transduced and then split onto mouse embryonic fibroblasts (MEFs). Cells were switched to HUES medium (KO-DMEM (Invitrogen), 10% KO-Serum Replacement (Invitrogen), 10% Plasmin (Talecris), 1× GlutaMAX (Invitrogen), 1× NEAA (Invitrogen), 1× 2-β-mercaptoethanol (Sigma) and 20 ng ml⁻¹ FGF2 (Invitrogen)) and 1 μg ml⁻¹ doxycycline (Sigma) was added to HUES medium for the first 21–28 days of reprogramming. hiPSCs were generally grown in HUES medium: early passage hiPSCs were split through manual passaging, whereas at higher passages hiPSCs could be enzymatically passaged with 1 mg ml⁻¹ collagenase (Sigma).

hiPSC differentiation to NPCs and neurons. Embryoid bodies were generated from hiPSCs and then transferred to non-adherent plates (Corning). Colonies were maintained in suspension in N2 medium (DMEM/F12 (Invitrogen), 1× N2 (Invitrogen)) for 7 days and then plated onto polyornithine (PORN)/laminin-coated plates. Visible rosettes formed within 1 week and were manually dissected and cultured in NPC medium (DMEM/F12, 1× N2, 1× B27-RA (Invitrogen), 1 μg ml⁻¹ laminin (Invitrogen) and 20 ng ml⁻¹ FGF2 (Invitrogen)). NPCs are maintained at high density, grown on PORN/laminin-coated plates in NPC medium and split approximately 1:4 every week with Accutase (Millipore). For neural differentiations, NPCs were dissociated with Accutase and plated at low density in neural differentiation medium (DMEM/F12-Glutamax, 1× N2, 1× B27-RA, 20 ng ml⁻¹ BDNF (Peprotech), 20 ng ml⁻¹ GDNF (Peprotech),

1 mM dibutyl-cyclic AMP (Sigma), 200 nM ascorbic acid (Sigma) onto PORN/laminin-coated plates. Assays for neuronal connectivity, neurite outgrowth, synaptic protein expression, synaptic density, electrophysiology, spontaneous calcium transient imaging and gene expression were used to compare control and SCZD hiPSC neurons.

1 mM dibutyl-cyclic AMP (Sigma), 200 nM ascorbic acid (Sigma) onto PORN/laminin-coated plates.

Assays for neuronal connectivity, neurite outgrowth, synaptic protein expression, synaptic density, electrophysiology, spontaneous calcium transient imaging and gene expression were used to compare control and SCZD hiPSC neurons.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 July 2010; accepted 10 February 2011.

Published online 13 April 2011.

- Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60**, 1187–1192 (2003).
- Wong, A. H. & Van Tol, H. H. Schizophrenia: from phenomenology to neurobiology. *Neurosci. Biobehav. Rev.* **27**, 269–306 (2003).
- Javitt, D. C., Spencer, K. M., Thaker, G. K., Winterer, G. & Hajos, M. Neurophysiological biomarkers for drug development in schizophrenia. *Nature Rev. Drug Discov.* **7**, 68–83 (2008).
- Ebert, A. D. et al. Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* **457**, 277–280 (2009).
- Lee, G. et al. Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* **461**, 402–406 (2009).
- Marchetto, M. C. et al. A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* **143**, 527–539 (2010).
- Maherali, N. et al. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55–70 (2007).
- Ugolini, G. Use of rabies virus as a transneuronal tracer of neuronal connections: implications for the understanding of rabies pathogenesis. *Dev. Biol. (Basel)* **131**, 493–506 (2008).

9. Wickersham, I. R. *et al.* Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53**, 639–647 (2007).
10. Lafon, M. Rabies virus receptors. *J. Neurovirol.* **11**, 82–87 (2005).
11. Selemon, L. D. & Goldman-Rakic, P. S. The reduced neuropil hypothesis: a circuit based model of schizophrenia. *Biol. Psychiatry* **45**, 17–25 (1999).
12. Jaaro-Peled, H., Ayhan, Y., Pletnikov, M. V. & Sawa, A. Review of pathological hallmarks of schizophrenia: comparison of genetic models with patients and nongenetic models. *Schizophr. Bull.* **36**, 301–313 (2010).
13. Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539–543 (2008).
14. Higgs, B. W., Elashoff, M., Richman, S. & Barci, B. An online database for brain disease research. *BMC Genomics* **7**, 70 (2006).
15. Patil, S. T. *et al.* Activation of mGlu2/3 receptors as a new approach to treat schizophrenia: a randomized Phase 2 clinical trial. *Nature Med.* **13**, 1102–1107 (2007).
16. Patterson, S. L. *et al.* Some forms of cAMP-mediated long-lasting potentiation are associated with release of BDNF and nuclear translocation of phospho-MAP kinase. *Neuron* **32**, 123–140 (2001).
17. Freyberg, Z., Ferrando, S. J. & Javitch, J. A. Roles of the Akt/GSK-3 and Wnt signaling pathways in schizophrenia and antipsychotic drug action. *Am. J. Psychiatry* **167**, 388–396 (2010).
18. Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).
19. The International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).
20. Karayiorgou, M. & Gogos, J. A. The molecular genetics of the 22q11-associated schizophrenia. *Brain Res. Mol. Brain Res.* **132**, 95–104 (2004).
21. Wang, X. D., Su, Y. A., Guo, C. M., Yang, Y. & Si, T. M. Chronic antipsychotic drug administration alters the expression of neuregulin 1 β , ErbB2, ErbB3, and ErbB4 in the rat prefrontal cortex and hippocampus. *Int. J. Neuropsychopharmacol.* **11**, 553–561 (2008).
22. Kapur, S. *et al.* PET evidence that loxapine is an equipotent blocker of 5-HT₂ and D₂ receptors: implications for the therapeutics of schizophrenia. *Am. J. Psychiatry* **154**, 1525–1529 (1997).
23. Cannon, T. D. & Keller, M. C. Endophenotypes in the genetic analyses of mental disorders. *Annu. Rev. Clin. Psychol.* **2**, 267–290 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements L. Moore, B. Miller, K. Stecker, J. Jepsen, D. Sepp, S. Larkin and L. Johnson provided technical assistance. T. Berggren directs, and M. Lutz manages, the Salk Stem Cell facility. D. Gibbs directs the Salk Viral Vector Core. J. Nguyen and L. Ouyang provided gene expression support. D. Chambers and J. Barrie provided FACS support. E. Callaway and I. Wickersham provided rabies trans-neuronal tracing viruses and invaluable advice and scientific feedback. M. Lawson provided assistance with statistical analysis. Thanks to G. Yeo, M. McConnell, S. Aigner, C. Marchetto and L. Boyer for advice and conversation. K.J.B. is supported by a training grant from the California Institute for Regenerative Medicine. The Gage Laboratory, and this project, is partially funded by CIRM Grant RL1-00649-1, The Lookout and Mathers Foundation, the Helmsley Foundation as well as Sanofi-Aventis.

Author Contributions K.J.B. designed the experiments with F.H.G. K.J.B. completed the experiments and wrote the manuscript. A.S. contributed to the microarray analysis and qPCR experiments. J.J. established the synaptic density assay and completed the calcium transient experiments. C.G.-B. and S.S. performed most of the synaptic protein experiments. N.T. analysed the rabies data. N.T. and S.S. together counted neurites. Y.L., Y.M. and G.C. performed electrophysiology. D.Y. established the calcium transient assay. S.M.C. and J.S. completed the CNV analysis.

Author Information The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE25673. As per our agreement with Coriell Cell Repository, all hiPSC lines generated from control and schizophrenic fibroblasts will only be available from Coriell. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.H.G. (gage@salk.edu).

METHODS

Description of SCZD patients. All patient samples were obtained from the Coriell collection. Patients were selected based on the high likelihood of a genetic component to disease. Patient 1 (GM02038, male, 22 years of age, Caucasian) was diagnosed with SCZD at 6 years of age and committed suicide at 22 years of age. Patient 2 (GM01792, male, 26 years of age, Jewish Caucasian) displayed episodes of agitation, delusions of persecution, and fear of assassination. His sister, patient 3 (GM01835, female, 27 years of age, Jewish Caucasian) had a history of schizoaffective disorder and drug abuse. Patient 4 (GM02497, male, 23 years of age, Jewish Caucasian) was diagnosed with SCZD at age 15 and showed symptoms including paralogical thinking, affective shielding, splitting of affect from content, and suspiciousness. His sister, patient 5 (GM02503, female, 27 years of age, Jewish Caucasian), was diagnosed with anorexia nervosa in adolescence and with schizoid personality disorder (SPD) as an adult. SPD has an increased prevalence in families with SCZD and is a milder diagnosis characterized not by psychosis, but rather by a lack of interest in social relationships and emotional coldness²⁴. Although we show data from SPD patient 5 as an interesting point of comparison, we do not consider patient 5 to belong to either the 'control' or 'SCZD' groups.

Preliminary experiments were controlled using BJ fibroblasts from ATCC (CRL-2522). These fibroblasts were expanded from foreskin tissue of a newborn male. They are readily reprogrammed, low passage, karyotypically normal and extremely well-characterized primary fibroblast line cells. Age- and ancestry-matched controls were obtained from three Coriell collections: apparently healthy individuals with normal psychiatric evaluations, apparently healthy non-fetal tissue and gerontology research centre cell cultures. hiPSCs were generated from GM02937 (male, 22 years of age), GM03440 (male, 20 years of age), GM03651 (female, 25 years of age), GM04506 (female, 22 years of age), AG09319 (female, 24 years of age) and AG09429 (female, 25 years of age).

Generation of lentivirus. Lentivirus was packaged in 293T HEK cells grown in 293T medium (IMEM (Invitrogen), 10% FBS (Gemini), 1× GlutaMAX (Invitrogen)). 293T cells were transfected with polyethylenimine (PEI) (Polysciences). Per 15-cm plate, the following solution was prepared, incubated for 5 min at room temperature and added drop-wise to plates: 12.2 µg lentiviral DNA, 8.1 µg MDL-gagpol, 3.1 µg Rev-RSV, 4.1 µg CMV-VSVG, 500 µl of IMDM and 110 µl PEI (1 µg µl⁻¹) and vortexed lightly. Medium was changed after 3 h and the virus was harvested at 48 and 72 h post transfection.

hiPSC derivation. Human fibroblasts were cultured on plates treated with 0.1% gelatin (in milli-Q water) for a minimum of 30 min and grown in HF medium (DMEM (Invitrogen), 10% FBS (Gemini), 1× GlutaMAX (Invitrogen), 5 ng ml⁻¹ FGF2 (Invitrogen)).

Human fibroblasts were infected daily for 5 days with tetracycline-inducible lentiviruses expressing *OCT4*, *SOX2*, *KLF4*, *cMYC* and *LIN28*, driven by a sixth lentivirus expressing the reverse tetracycline transactivator (rtTA)⁷. Cells from a single well of a 6-well dish were split onto a 10-cm plate containing 10⁶ mouse embryonic fibroblasts (mEFs). Cells were switched to HUES medium (KO-DMEM (Invitrogen), 10% KO-Serum Replacement (Invitrogen), 10% Plasmanate (Talecris), 1× GlutaMAX (Invitrogen), 1× NEAA (Invitrogen), 1× 2-β-mercaptoethanol (Sigma) and 20 ng ml⁻¹ FGF2 (Invitrogen)). Doxycycline (1 µg ml⁻¹; Sigma) was added to HUES medium for the first 21–28 days of reprogramming.

hiPSC colonies were picked manually and clonally plated onto 24-well mEF plates. hiPSC lines were either maintained on mEFs in HUES medium or on Matrigel (BD) in TeSR medium (StemCell Technologies). At early passages, hiPSCs were split through manual passaging. At higher passages, hiPSC could be enzymatically passaged with collagenase (1 mg ml⁻¹ in DMEM; Sigma). Cells were frozen in freezing medium (DMEM, 10% FBS, 10% DMSO).

Karyotyping analysis was performed by Cell Line Genetics or by M. Dell'Aquila. Teratoma analysis was performed by injecting hiPSCs into the kidney capsules of isoflurane-anesthetized NOD-SCID mice. Teratomas were harvested eight weeks post-injection, paraffin-embedded and stained with haematoxylin and eosin.

hiPSC differentiation to NPCs and neurons. hiPSCs grown in HUES medium on mEFs were incubated with collagenase (1 mg ml⁻¹ in DMEM) at 37 °C for 1–2 h until colonies lifted from the plate and were transferred to a non-adherent plate (Corning). Embryoid bodies were grown in suspension in N2 medium (DMEM/F12-GlutaMAX (Invitrogen), 1× N2 (Invitrogen)). After 7 days, embryoid bodies were plated in N2 medium with 1 µg ml⁻¹ laminin (Invitrogen) onto polyornithine (PORN)/laminin-coated plates. Visible rosettes formed within one week and were manually dissected onto PORN/laminin-coated plates. Rosettes were cultured in NPC medium (DMEM/F12, 1× N2, 1× B27-RA (Invitrogen), 1 µg ml⁻¹ Laminin and 20 ng ml⁻¹ FGF2) and dissociated in TrypLE (Invitrogen) for 3 min at 37 °C. NPCs are maintained at high density, grown on PORN/laminin-coated plates in NPC medium and split approximately 1:4 every week with Accutase (Millipore).

For neural differentiations, NPCs were dissociated with Accutase and plated in neural differentiation medium (DMEM/F12, 1× N2, 1× B27-RA, 20 ng ml⁻¹ BDNF (Peprotech), 20 ng ml⁻¹ GDNF (Peprotech), 1 mM dibutyryl-cyclic AMP (Sigma), 200 nM ascorbic acid (Sigma) onto PORN/laminin-coated plates. Density is critical and the following guidelines were used: 2-well Permanox slide, 80,000–100,000 cells per well; 24-well, 40,000–60,000 cells per well; 6-well, 200,000 cells per well. hiPSC-derived neurons were differentiated for 1–3 months. Notably, synapse maturation occurs most robustly *in vitro* when hiPSC neurons are cultured together with wild-type human cerebellar astrocytes (ScienCell). FBS (0.5%) was supplemented into neural differentiation media for all astrocyte coculture experiments.

It is difficult to maintain healthy neurons for 3 months of differentiation and some cultures invariably fail or become contaminated. When even one SCZD patient neural culture failed, the experiments were abandoned as all assays were conducted on neurons cultured in parallel. If, however, only a control neural culture failed, and at least three control samples remained, analysis was completed. For this reason, although patients are consistently numbered throughout the manuscript, controls are not, and are instead listed in numerical order (BJ, GM02937, GM03651, GM04506, AG09319, AG09429).

Antipsychotic drugs were added for the final 3 weeks of a 3-month differentiation on astrocytes and for the final two weeks of a 6-week differentiation on PORN/laminin alone. Drugs were resuspended in DMSO at the following concentrations: clozapine (5 µM), loxapine (10 µM), olanzapine (1 µM), risperidone (10 µM) and thioridazine (5 µM).

Immunohistochemistry. Cells were fixed in 4% paraformaldehyde in PBS at 4 °C for 10 min. hiPSCs and NPCs were permeabilized at room temperature for 15 min in 1.0% Triton in PBS. All cells were blocked in 5% donkey serum with 0.1% Triton at room temperature for 30 min. The following primary antibodies and dilutions were used: mouse anti-Oct4 (Santa Cruz), 1:200; goat anti-Sox2 (Santa Cruz), 1:200; goat anti-Nanog (R&D), 1:200; mouse anti-Tral-60 (Chemicon), 1:100; mouse anti-human Nestin (Chemicon), 1:200; rabbit anti-βIII-tubulin (Covance), 1:200; mouse anti-βIII-tubulin (Covance), 1:200; rabbit anti-cow-GFAP (Dako) 1:200; mouse anti-MAP2AB (Sigma), 1:200; rabbit anti-synapsin (Synaptic Systems), 1:500; mouse anti-PSD95 (UCDavis/NIH Neuromab), 1:500; rabbit anti-PSD95 (Invitrogen), 1:200; rabbit anti-VGLUT1 (Synaptic Systems), 1:500; rabbit anti-gephyrin, (Synaptic Systems), 1:500; mouse anti-vGAT (Synaptic Systems), 1:500; rabbit anti-vGat (Synaptic Systems), 1:500; rabbit anti-GLUR1 (Oncogene), 1:100; rabbit anti-GABA (Sigma), 1:200; rabbit anti-GAD65/67 (Sigma), 1:200.

Secondary antibodies were Alexa donkey 488, 555 and 647 anti-rabbit (Invitrogen), Alexa donkey 488 and 555 anti-mouse (Invitrogen), and Alexa donkey 488, 555, 568 and 594 anti-goat (Invitrogen); all were used at 1:300. To visualize nuclei, slides were stained with 0.5 µg ml⁻¹ DAPI (4',6-diamidino-2-phenylindole) and then mounted with Vectashield. Images were acquired using a Bio-Rad confocal microscope.

FACS. For sorting of dissociated two-month-old hiPSC neurons, cultures were dissociated with Accutase for 5 min, washed in DMEM, centrifuged at 500g and resuspended in PBS. Cells were fixed in 4% paraformaldehyde in PBS at 4 °C for 10 min. Cells were washed in PBS and aliquoted into 96-well conical plates. Cells were blocked in 5% donkey serum with 0.1% saponin at room temperature for 30 min. The following primary antibodies and dilutions were used for one hour at room temperature: rabbit anti-βIII-tubulin (Sigma), 1:200; mouse anti-βIII-tubulin (Covance), 1:200; rabbit anti-GAD65/67 (Sigma), 1:200. Cells were washed and then incubated with secondary antibodies at 1:200 for 30 min at room temperature: Alexa donkey 647 anti-rabbit (Invitrogen), and Alexa donkey 488 anti-mouse (Invitrogen). Cells were washed three times in PBS and stained with 0.5 µg ml⁻¹ DAPI. Cells were resuspended in PBS with 5% donkey serum and 0.1% detergent saponin. The homogeneous solution was filtered through a 250-µm nylon sieve and run in a BD FACSCalibur. Data were analysed using FloJo.

Rabies virus trans-neuronal tracing. Rabies virus trans-neuronal tracing was performed on 3-month-old hiPSC neurons grown together with wild-type human astrocytes (ScienCell) on acid-etched glass coverslips and then transduced with LV-SYNP-HTG or LV-SYNP-HT. Cultures were transduced with Rabies-ENVAΔG-RFP after at least a week to allow expression of ENVA and rabies G. Either 5, 7 or 10 days later, hiPSC neurons were fixed with 4% paraformaldehyde in PBS for fluorescent microscopy or FACS analysis; cells for FACS analysis were first dissociated with Accutase before fixation.

Neurite analysis. Neurite analysis was performed on 3-month-old hiPSC neurons grown together with wild-type human astrocytes (ScienCell) on acid-etched glass coverslips. Low titre transduction of a lentivirus driving expression of GFP from the SYN promoter (LV-SYNP-GFP) occurred at least 7 days before assay. LV-SYNP-GFP was used to image and count branching neurites from single neurons

(Fig. 3a). The number of neurites extending from the soma of 691 single LV-SYNP-GFP-labelled neurons was determined by a blinded count.

Synaptic protein staining analysis. Synaptic protein staining was performed on 3-month-old hiPSC neurons grown together with wild-type human astrocytes (ScienCell) on acid-etched glass coverslips. To calculate ratios of MAP2AB-positive dendrites and synaptic proteins, confocal images were taken at $\times 630$ magnification and $\times 4$ zoom. Using NIH ImageJ, images were thresholded and the integrated pixel density was determined for each image. Integrated pixel density measurement is the product of area (measured in square pixels) and mean grey value (the sum of the grey values of all the pixels in the selection divided by the number of pixels).

Synapse density. Synapse density analysis was performed on 3-month-old hiPSC neurons grown together with wild-type human astrocytes (ScienCell) on acid-etched glass coverslips. Manual counts of synaptic density were done in three steps using NIH ImageJ. First, the colocalization plugin was used to identify colocalization of VGLUT1 and PSD95. Second, the particle analysis function was used to restrict size 50-infinity. Third, dendrites were traced using the NeuronJ plugin. The mask generated by particle analysis was overlaid on the trace generated by NeuronJ and synapses were manually counted.

Electrophysiology. Whole-cell perforated patch recordings were performed on SCZD ($n = 30$) and control ($n = 20$) 3-month-old hiPSC neurons grown together with wild-type human astrocytes (ScienCell) on acid-etched coverslips and typically transduced with LV-SYNP-GFP. The recording micropipettes (tip resistance 3–6 M Ω) were tip-filled with internal solution composed of 115 mM K-gluconate, 4 mM NaCl, 1.5 mM MgCl₂, 20 mM HEPES, and 0.5 mM EGTA (pH 7.4) and then back-filled with the same internal solution containing 200 μ g ml⁻¹ amphotericin B (Calbiochem). Recordings were made using an Axopatch 200B amplifier (Axon Instruments). Signals were sampled and filtered at 10 kHz and 2 kHz, respectively. The whole-cell capacitance was fully compensated, whereas the series resistance was uncompensated but monitored during the experiment by the amplitude of the capacitive current in response to a 5-mV pulse. The bath was constantly perfused with fresh HEPES-buffered saline composed of 115 mM NaCl, 2 mM KCl, 10 mM HEPES, 3 mM CaCl₂, 10 mM glucose and 1.5 mM MgCl₂ (pH 7.4). For voltage-clamp recordings, cells were clamped at -60 to -80 mV; Na⁺ currents and K⁺ currents were stimulated by voltage step depolarizations. Command voltage varied from -50 to $+20$ mV in 10 mV increments. For current-clamp recordings, induced action potentials were stimulated with current steps from -0.2 to $+0.5$ nA. All recordings were performed at room temperature.

Spontaneous calcium transients. Calcium imaging analysis was performed on 2.5- to 3-month-old hiPSC neurons grown together with wild-type human astrocytes (ScienCell) on acid-etched glass coverslips. Culture medium was removed and hiPSC cultures were incubated with 0.4 μ M Fluo-4AM (Molecular Probes) and 0.02% Pluronic F 127 detergent in Krebs HEPES buffer (KHB) (10 mM HEPES, 4.2 mM NaHCO₃, 10 mM dextrose, 1.18 mM MgSO₄·2H₂O, 1.18 mM KH₂PO₄, 4.69 mM KCl, 118 mM NaCl, 1.29 mM CaCl₂; pH 7.3) for 1 h at room temperature. Cells were washed with KHB buffer, incubated for 2 min with Hoechst dye diluted 1:1,000 in KHB, and allowed to incubate for an additional 15 min in KHB to equilibrate intracellular dye concentration. Time-lapse image sequences ($\times 100$ magnification) were acquired at 28 Hz using a Hamamatsu ORCA-ER digital camera with a 488-nm (FITC) filter on an Olympus IZ81 inverted fluorescence confocal microscope. Images were acquired with MetaMorph.

In total, eight independent neural differentiations were tested per patient, 210 movies of spontaneous calcium transients (110 control and 100 SCZD) were generated and 2,676 regions of interest (1,158 control and 1,518 SCZD) were analysed. Up to four 90-s videos of Fluo-4AM fluorescence were recorded per neural differentiation per patient with a spinning disc confocal microscope at 28 frames per second (Supplementary Fig. 5A). Using ImageJ software, regions of interest can be manually selected and the mean pixel intensity of each region of interest can be followed over time, generating time trace data for each region of interest. The data were analysed in Matlab where background subtraction was performed by normalizing traces among traces of the sample, and spike events were identified based on the slope and amplitude of the time trace.

The amplitude of spontaneous calcium transients was calculated by measuring the change in total pixel intensity for each normalized calcium transient trace. The

rate was determined by dividing the total number of spontaneous calcium transients for any regions of interest by the total length of the movie (90 s). The synchronicity of spontaneous calcium transients was determined by two independent calculations. First, to determine the percentage synchronicity per calcium transient, the total number of synchronized calcium transients, defined as three or more simultaneous peaks, was divided by the total number of spontaneous calcium transients identified. Second, to calculate the maximum percentage synchronicity, the maximum number of regions of interest involved in a single synchronized event was divided by the total number of regions of interest identified.

CNV analysis. Cells were lysed in DNA lysis solution (100 mM Tris, pH 8.5, 5 mM EDTA, 200 mM NaCl, 0.2% (w/v) sarcosyl and 100 μ g ml⁻¹ fresh proteinase K) overnight at 50 °C. DNA was precipitated by the addition of an equal volume of NaCl-ethanol mixture (150 μ l of 5 M NaCl in 10 ml cold 95% ethanol) and then washed three times in 70% ethanol before resuspension in water with RNase A overnight at 4 °C.

Genome scans were performed using NimbleGen HD2 arrays (NimbleGen Systems) according to the manufacturer's instructions using a standard reference genome SKNI. NimbleGen HD2 dual-colour intensity data were normalized in a two-step process: first, a 'spatial' normalization of probes was performed to adjust for regional differences in intensities across the surface of the array; second, the Cy5 and Cy3 intensities were adjusted to a fitting curve by invariant set normalization, preserving the variability in the data. The log₂ ratio for each probe was then estimated using the geometric mean of normalized and raw intensity data²⁵.

CNV analysis was completed to identify deletions and duplications present within our patients. By using a virtual 'genotyping' step whereby individual CNV segment probe ratios were converted into z-scores, a distribution of median z-scores was generated, outliers of which were considered to be true CNVs. In doing so, we better filtered out common artefacts and false-positive CNVs and generated a list of CNVs unbiased by previous genetic studies of SCZD.

Patient fibroblasts were used for CNV analysis. Lymphocytes were available for patients 4 and 5 and their parents, allowing us to validate the CNVs identified for patient 4 and also to determine the parent of origin for each mutation; many were inherited from the unaffected mother (Supplementary Table 7).

Gene expression analysis. Unless otherwise specified, gene expression analysis was performed on 6-week-old hiPSC neurons without astrocyte coculture. Cells were lysed in RNA BEE (Tel-test). RNA was chloroform-extracted, pelleted with isopropanol, washed with 70% ethanol and resuspended in water. RNA was treated with RQ1 RNase-free DNase (Promega) for 30 min at 37 °C and then the reaction was inactivated by incubation with EGTA stop buffer at 65 °C for 10 min.

For gene expression microarrays, three independent neural differentiations for each of the four SCZD patients as well as four control subjects were compared using Affymetrix Human 1.0ST arrays as specified by the manufacturer.

Gene expression microarray analysis was completed using Partek Genomics Suite software. Intensity values were generated as follows: RMA background correction, quantile normalization, log₂ transformation and mean polished probeset summarization. Pathway analysis was performed using Metacore GeneGo.

For qPCR, cDNA was synthesized using Superscript III at 50 °C for 1–2 h, inactivated for 15 min at 70 °C and then treated with RNase H for 15 min at 37 °C, inactivated with EDTA and heated to 70 °C for 15 min. qPCR was performed using SYBR Green. Primers used are listed in Supplementary Table 8.

Statistical analysis. Statistical analysis was performed using JMP. Box-Cox transformation of raw data was performed to correct non-normal distribution of the data and residuals. Improvements were assessed by Shapiro-Wilk *W* test of the transformed data and residuals. Means were compared within diagnosis by one-way analysis using both Student's *t*-test and Tukey-Kramer HSD. Finally, a nested analysis of values for individual patients was performed using standard least squares analysis comparing means for all pairs using Student's *t*-test for specific pairs and Tukey-Kramer HSD for multiple comparisons.

24. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, fourth edition (DSM-IV)*. (American Psychiatric Press, 1994).
25. McCarthy, S. E. et al. Microduplications of 16p11.2 are associated with schizophrenia. *Nature Genet.* **41**, 1223–1227 (2009).

N-acylethanolamine signalling mediates the effect of diet on lifespan in *Caenorhabditis elegans*

Mark Lucanic¹, Jason M. Held¹, Maithili C. Vantipalli¹, Ida M. Klang^{1,2}, Jill B. Graham¹, Bradford W. Gibson¹, Gordon J. Lithgow¹ & Matthew S. Gill^{1†}

Dietary restriction is a robust means of extending adult lifespan and postponing age-related disease in many species, including yeast, nematode worms, flies and rodents^{1,2}. Studies of the genetic requirements for lifespan extension by dietary restriction in the nematode *Caenorhabditis elegans* have implicated a number of key molecules in this process^{3–5}, including the nutrient-sensing target of rapamycin (TOR) pathway⁶ and the Foxa transcription factor PHA-4 (ref. 7). However, little is known about the metabolic signals that coordinate the organismal response to dietary restriction and maintain homeostasis when nutrients are limited. The endocannabinoid system is an excellent candidate for such a role given its involvement in regulating nutrient intake and energy balance⁸. Despite this, a direct role for endocannabinoid signalling in dietary restriction or lifespan determination has yet to be demonstrated, in part due to the apparent absence of endocannabinoid signalling pathways in model organisms that are amenable to lifespan analysis⁹. N-acylethanolamines (NAEs) are lipid-derived signalling molecules, which include the mammalian endocannabinoid arachidonoyl ethanolamide. Here we identify NAEs in *C. elegans*, show that NAE abundance is reduced under dietary restriction and that NAE deficiency is sufficient to extend lifespan through a dietary restriction mechanism requiring PHA-4. Conversely, dietary supplementation with the nematode NAE eicosapentaenoyl ethanolamide not only inhibits dietary-restriction-induced lifespan extension in wild-type worms, but also suppresses lifespan extension in a TOR pathway mutant. This demonstrates a role for NAE signalling in ageing and indicates that NAEs represent a signal that coordinates nutrient status with metabolic changes that ultimately determine lifespan.

We identified a diverse set of NAEs in *C. elegans* using stable isotope dilution gas chromatography mass spectrometry (SID-GC-MS)¹⁰ in pseudo-multiple reaction monitoring (pMRM) mode (Fig. 1a and Supplementary Fig. 1), including the C20-fatty-acid-containing NAEs eicosapentaenoyl ethanolamide (EPEA) and arachidonoyl ethanolamide (AEA), the latter having been previously identified in *C. elegans*¹¹. Mammalian NAE levels are controlled through enzymatic synthesis and degradation, whereas their biological effects are mediated through interactions with several receptors, including the cannabinoid receptors that bind AEA⁸. Many of the upstream enzymes that regulate NAEs remain unidentified; however, N-acyl-phosphatidylethanolamine-specific phospholipase D (NAPE-PLD) catalyses the last step in biosynthesis, whereas the hydrolytic enzyme fatty acid amide hydrolase (FAAH) inactivates NAE molecules¹². Although *C. elegans* lacks clear orthologues of cannabinoid receptors⁹, it does have orthologues of NAPE-PLD (*nape-1*) and FAAH (*faah-1*)¹³. *nape-1* is expressed in interneurons that are in close proximity to the primary sensory neurons, and *faah-1* is expressed principally in the pharynx, indicating that this is a major site of NAE degradation (Supplementary Fig. 2).

We hypothesized that if *C. elegans nape-1* and *faah-1* have conserved function then perturbation of their activity should alter worm

NAE levels. Reducing the levels of *faah-1* through RNA interference (RNAi) or inhibiting enzyme activity with a specific chemical inhibitor increased levels of NAEs (Fig. 1b, c and Supplementary Figs 3 and 4). In contrast, no decrease in NAE levels was observed in either a *nape-1* deletion mutant or after RNAi of *nape-1* in wild-type animals (data not shown), which is consistent with mammalian studies that indicate redundancy at the final step of NAE production¹⁴. We then reasoned that overexpression of *faah-1* would provide an alternative means of reducing NAE levels *in vivo*, and found that transgenic worm strains maintaining extra copies of the *faah-1* gene had reduced EPEA levels, as well as reductions in palmitoleoyl ethanolamide, linoleoyl ethanolamide and AEA (Fig. 1d and Supplementary Fig. 5). These data indicate that the function of *faah-1* as a hydrolytic enzyme involved in NAE degradation is conserved in worms.

A marked phenotype of *faah-1* overexpression in worms was a developmental delay which could be rescued by *faah-1* RNAi (Fig. 2a, b), indicating that NAEs promote larval development. During normal growth all NAEs have a similar developmental profile, reaching the

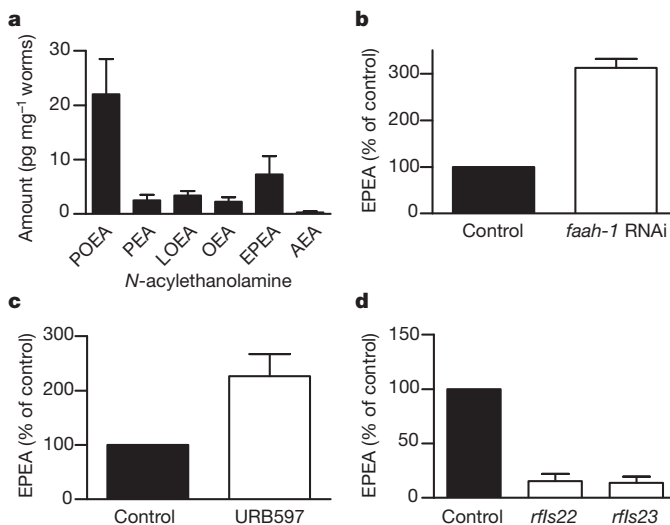


Figure 1 | NAE levels in *C. elegans* are modulated by FAAH activity. **a**, Levels of NAEs in first-day adult wild-type N2 worms measured by SID-GC-MS (mean + s.d., $n = 5$). AEA, arachidonoyl ethanolamide; EPEA, eicosapentaenoyl ethanolamide; LOEA, linoleoyl ethanolamide; OEA, oleoyl ethanolamide; PEA, palmitoyl ethanolamide; POEA, palmitoleoyl ethanolamide. **b**, EPEA levels are elevated in first-day *eri-1(mg366); lin-15B(n744)* adults after exposure to *faah-1* dsRNA by soaking (mean + s.d., $n = 2$). **c**, EPEA levels are elevated in first-day wild-type N2 adults after 24 h exposure to 10 μ M URB597, a chemical inhibitor of mammalian FAAH (mean + s.d., $n = 5$, $P < 0.05$, Wilcoxon signed rank test). **d**, Overexpression of *faah-1* results in reduced EPEA levels in first-day wild-type N2 adults (mean + s.d.; N2, $n = 9$; *rfls22*, $n = 7$; and *rfls23*, $n = 8$, $P < 0.05$ for both *rfls22* and *rfls23*, Wilcoxon signed rank test).

¹Buck Institute for Research on Aging, 8001 Redwood Boulevard, Novato, California 94945, USA. ²Karolinska Institute, Center for Biosciences at NOVUM, Department of Biosciences and Nutrition, Hälsovägen 7, S-141 83 Huddinge, Sweden. [†]Present address: The Scripps Research Institute – Scripps Florida, 130 Scripps Way, 3B3, Jupiter, Florida 33458, USA.

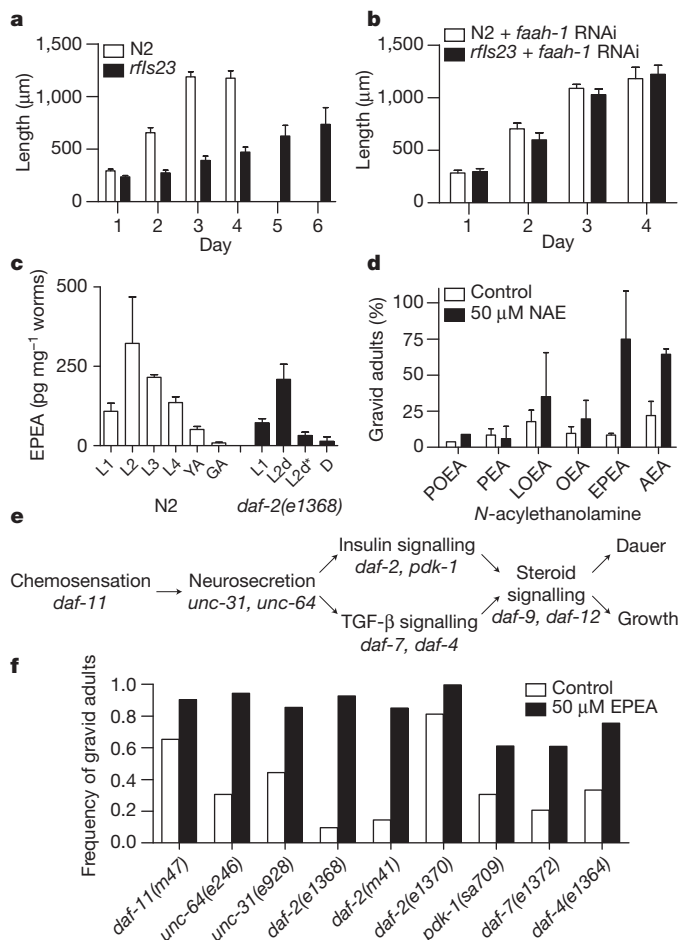


Figure 2 | NAEs affect reproductive growth and dauer formation. **a**, *faah-1* overexpression results in developmental delay (mean + s.d.; N2, $n = 54$; *rfls23*, $n = 76$). **b**, *faah-1* RNAi rescues the growth delay of *faah-1* overexpressors (mean + s.d.; N2, $n = 59$; *rfls23*, $n = 53$). **c**, Levels of EPEA during development in N2 and *daf-2(e1368)* animals grown at 25 °C (mean + s.d., $n = 2$). D, dauer; GA, gravid adult; L1, first larval stage; L2, second larval stage; L2d, alternative L2 stage preceding the dauer moult; L2d*, later time point in L2d; L3, third larval stage; L4, fourth larval stage; YA, young adult. **d**, Effect of treatment with exogenous NAEs on reproductive growth in *daf-2(e1368)* mutants at 24 °C (mean + s.d., $n = 2$). **e**, Scheme illustrating genes and pathways involved in dauer formation in *C. elegans*. **f**, EPEA rescues dauer formation in multiple dauer constitutive mutants (all $P < 0.0001$, chi-squared test, additional data in Supplementary Table 1).

highest levels in the second larval stage (L2) and then progressively declining into adulthood (Fig. 2c and Supplementary Fig. 6). The peak in NAE levels at L2 coincides with the time at which the animal is committed to reproductive growth rather than entry into an alternative diapause stage (the dauer larva, Supplementary Fig. 6)¹⁵. Entry into the dauer state allows the worm to survive for extended periods of time in the absence of food and is associated with profound metabolic changes. We found that NAE levels under dauer-inducing conditions were similar to normally developing worms in L1 and early pre-dauer (L2d) but were markedly reduced during the late L2d and dauer stages (Fig. 2c and Supplementary Fig. 6). The reduction in NAE levels after commitment to dauer formation indicates that NAEs could act as signals of an altered metabolic state.

Because NAE levels are reduced in worms during the transition to the dauer state, we hypothesized that exogenous NAEs could prevent dauer arrest and promote reproductive growth. Only EPEA was able to rescue completely the dauer phenotype of *daf-2(e1368)* mutants, whereas the less abundant AEA had a weaker effect (Fig. 2d and Supplementary Fig. 7). Importantly, EPEA concentrations in adult

animals treated with 50 μM EPEA reached a similar level to those found in wild-type L2 larvae (data not shown), indicating that exogenous EPEA treatment results in physiological levels *in vivo*. Treatment with eicosapentaenoic acid, which is both a precursor and the hydrolytic breakdown product of EPEA, was unable to promote reproductive growth, indicating that EPEA itself is responsible for the dauer rescue (Supplementary Fig. 8). EPEA was also able to rescue dauer formation in other dauer constitutive mutants that define distinct signalling pathways involved in dauer formation (Fig. 2e, f and Supplementary Table 1), indicating that this molecule functions downstream or parallel to these primary environment sensing pathways. Several candidate NAE receptors were not required for rescue of the dauer phenotype (Supplementary Fig. 9).

These data suggest that NAEs could provide a signal of nutrient availability and energy balance. In support of this, we found that starved L1 animals showed extremely low levels of EPEA and the other NAEs, all of which increased markedly after 6 h of feeding (Fig. 3a and Supplementary Fig. 10). Moreover, NAE levels were reduced in adult animals maintained under dietary restriction conditions, and re-feeding restored NAE levels back to that of well-fed controls (Fig. 3b and Supplementary Fig. 11). These results demonstrate that NAE levels in worms are responsive to nutrient availability, as has been shown in rodent studies^{16,17}.

Because dietary restriction reduces NAEs, we hypothesized that *faah-1* overexpression may partly mimic a dietary restriction state and confer stress resistance and longevity phenotypes. Consistent with this hypothesis, we found that worms overexpressing *faah-1* showed resistance to thermal stress (Supplementary Table 2) and increased adult lifespan (Fig. 3c and Supplementary Table 3). Overexpression of *faah-1* in the pharynx was largely sufficient for this lifespan extension (Supplementary Table 3), supporting the idea that the pharynx is a major site of NAE function.

Using a standard dietary restriction paradigm¹⁸ we found that *faah-1* overexpression was associated with lifespan extension in the presence of abundant food but not under conditions of optimal dietary restriction (Fig. 3c–f and Supplementary Table 3). This was confirmed using a second, independent dietary restriction protocol¹⁹ (Supplementary Table 3). The lack of an additive effect under optimal dietary restriction conditions indicates that lifespan extension resulting from *faah-1* overexpression is mechanistically equivalent to dietary restriction. This is further supported by the observations of growth delay and reduced fecundity in animals with reduced NAEs (Fig. 2a and Supplementary Fig. 12). Lifespan extension resulting from *faah-1* overexpression was not dependent on DAF-16, the FOXO transcription factor required for longevity in *daf-2* insulin-signalling mutants (Fig. 3g and Supplementary Table 3). However, the transcription factor *pha-4*, which is involved in the lifespan-extending effects of dietary restriction⁷, was required (Fig. 3h and Supplementary Table 3), indicating that low NAE levels extend lifespan through a dietary restriction pathway and act upstream of PHA-4.

Because reduced NAE levels are associated with increased longevity, we hypothesized that elevated NAE levels should suppress stress resistance and lifespan extension. Consistent with this hypothesis, EPEA treatment resulted in a significant reduction in thermotolerance (Supplementary Table 2) and lifespan (Fig. 4a, b and Supplementary Table 4). Lifespan suppression induced by EPEA was minimal under conditions of high food but was much more profound under optimal dietary restriction conditions (Fig. 4c, d and Supplementary Table 4). This is unlikely to be due to EPEA toxicity, as EPEA treatment does not adversely affect growth (data not shown), has no effect on fertility (Supplementary Fig. 13), and only has a minimal effect on lifespan under high food conditions. Instead, we propose that elevated EPEA levels under dietary restriction provide a false signal of high nutrient availability and inhibit the metabolic adaptation to reduced food availability that confers lifespan extension.

We next sought to determine whether mutations that affect either NAE biosynthesis or nutrient sensing pathways also alter NAE levels

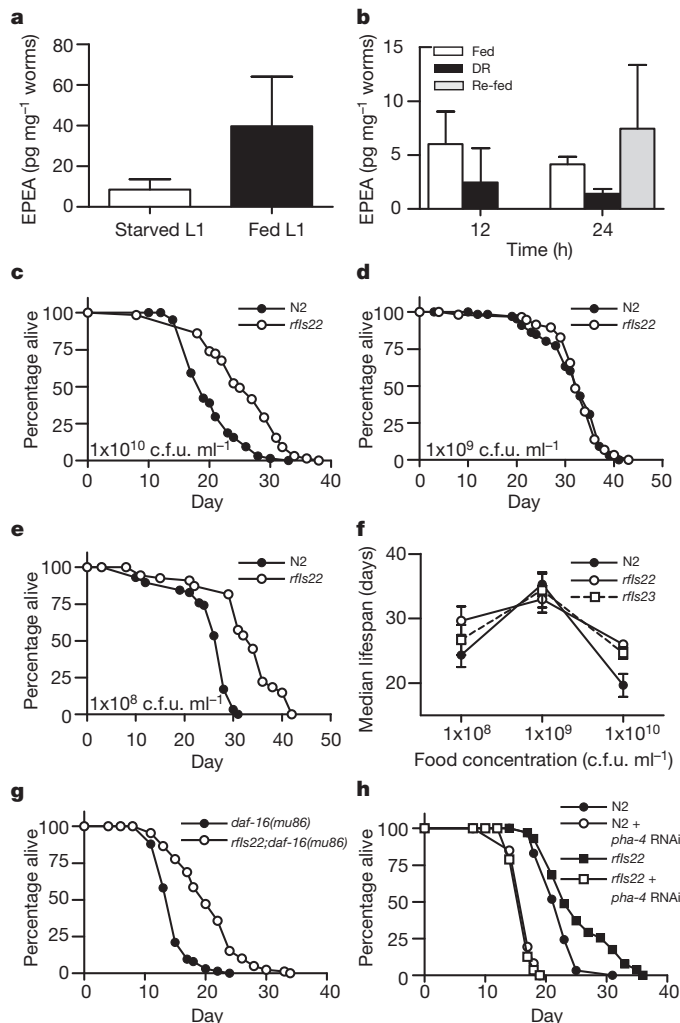


Figure 3 | Reduced NAE levels are associated with dietary restriction and are sufficient to confer lifespan extension. **a**, EPEA levels are reduced in starved L1 larvae and increase after 6 h of exposure to food (mean \pm s.d., $n = 3$). **b**, EPEA levels are altered in response to food availability in adult wild-type N2 animals (mean \pm s.d., Mann-Whitney U -test: 12 h fed ($n = 6$) versus dietary restriction (DR; $n = 12$), $P < 0.05$; 24 h fed ($n = 7$) versus dietary restriction ($n = 7$), $P < 0.001$; 24 h dietary restriction versus re-fed ($n = 6$), $P < 0.005$; 24 h fed versus re-fed, P = not significant). **c**, *faah-1* overexpression extends lifespan in N2 wild-type animals under fed conditions (1×10^{10} colony-forming units (c.f.u.) ml^{-1} *Escherichia coli*, $P < 0.0001$, log-rank test). **d**, Lifespan is not different between N2 and a *faah-1* overexpressing line under conditions of optimal dietary restriction (1×10^9 c.f.u. ml^{-1} *E. coli*). **e**, *faah-1* overexpression extends lifespan in N2 wild-type animals under conditions of sub-optimal dietary restriction conditions (1×10^8 c.f.u. ml^{-1} *E. coli*, $P < 0.0001$, log-rank test). **f**, *faah-1* overexpression affects lifespan in a nutrient-dependent manner (mean lifespan \pm s.d., $n = 3$). **g**, *faah-1* overexpression extends lifespan in a *daf-16* mutant ($P < 0.0001$, log-rank test). **h**, Lifespan extension resulting from *faah-1* overexpression requires the Foxa transcription factor PHA-4 (N2 control versus N2 plus *pha-4* RNAi, $P < 0.0001$; *rfls22* control versus *rfls22* plus *pha-4* RNAi, $P < 0.0001$; *rfls22* control versus N2 control, $P = 0.0014$; log-rank test).

and lifespan. In rodents, the dietary availability of fatty acids has been shown to influence NAE levels²⁰, and thus reduced fatty acid availability could generate NAE deficiency. In *C. elegans*, mutations in *fat-4* lead to reduced arachidonic and eicosapentaenoic acid²¹ and also result in reduced AEA and EPEA levels (Supplementary Fig. 14) as well as lifespan extension (Supplementary Fig. 15 and Supplementary Table 5). In addition, we found that *rsk-1* mutants, which have a defect in the worm orthologue of S6 kinase in the conserved TOR nutrient sensing pathway⁶, showed a specific reduction of EPEA levels, but not the other NAEs (Fig. 4e and Supplementary Fig. 16). Furthermore,

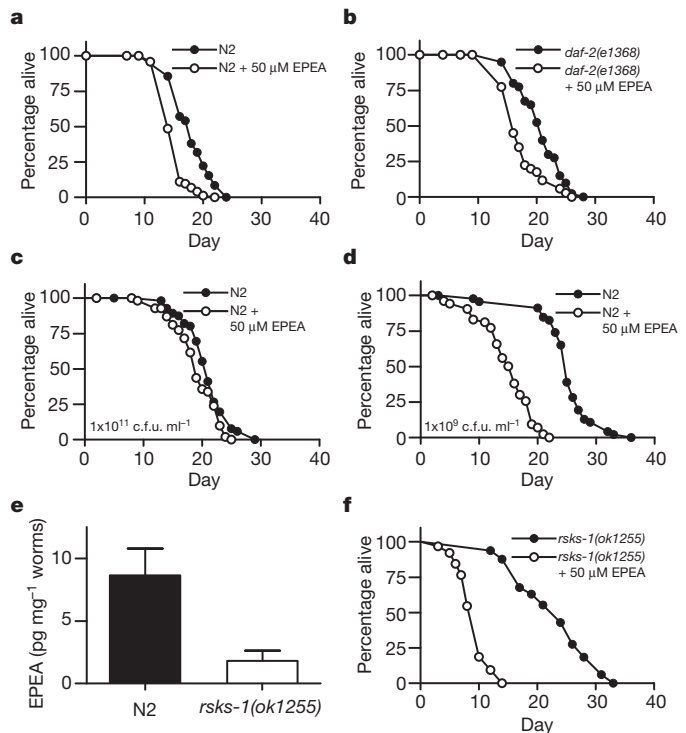


Figure 4 | EPEA suppresses the effects of dietary restriction on lifespan. **a**, EPEA treatment reduces lifespan in wild-type N2 animals on control RNAi bacteria ($P < 0.0001$, log-rank test). **b**, EPEA treatment reduces lifespan in *daf-2(e1368)* mutants on control RNAi bacteria ($P = 0.0005$, log-rank test). **c**, EPEA has a minimal effect on N2 lifespan in the presence of high food concentrations (1×10^{11} c.f.u. ml^{-1} *E. coli*, $P < 0.0001$, log-rank test). **d**, EPEA treatment completely suppresses the effect of optimal dietary restriction on wild-type N2 lifespan (1×10^9 c.f.u. ml^{-1} *E. coli*, $P < 0.0001$; log-rank test). **e**, EPEA levels are reduced in *rsk-1(ok1255)* mutants, a genetic model of dietary restriction (mean \pm s.d., $n = 4$, $P < 0.05$, Mann-Whitney U -test). **f**, EPEA treatment suppresses lifespan extension in *rsk-1(ok1255)* mutants ($P < 0.0001$, log-rank test).

EPEA treatment in *rsk-1* mutants completely suppressed their longevity phenotype (Fig. 4f and Supplementary Table 6). These data indicate that the TOR pathway may control NAE levels in response to nutrient availability and that the longevity of *rsk-1* mutants could be due to their inability to upregulate EPEA in response to food.

NAEs have emerged as an important class of lipid mediators with a role in the response to nutrient availability in diverse organisms including mammals, non-mammalian vertebrates and invertebrates^{16,17,22–25}. In mammals, AEA, an arachidonic acid containing NAE, elicits many of its effects through cannabinoid receptors, but can also interact with a variety of other targets. Although nematode worms possess a number of different NAEs, including AEA and EPEA, *C. elegans*, in common with other protostomes and some primitive deuterostomes^{26,27}, does not possess clear orthologues of the mammalian cannabinoid receptors⁹. This suggests that there are unidentified NAE receptors in nematodes that are possibly conserved mediators of NAE signalling. Taken together, our findings indicate that reduced NAE signalling mediates some of the effects of dietary restriction on lifespan extension in the nematode, and that EPEA acts as a metabolic signal that couples nutrient availability with growth and lifespan, suggesting a new role for NAE signalling in organismal ageing.

METHODS SUMMARY

Caenorhabditis elegans strains used in this study are described in Methods and were maintained as previously described²⁸. Generation of transgenic strains is described in Methods. The growth of worms in mass culture, lipid extractions and GC-MS are described in Methods. Dauer assays, thermotolerance assays and lifespan analysis were performed as previously described^{29,30} with exceptions

detailed in Methods. Dietary restriction experiments were performed according to the method of ref. 18.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 March 2010; accepted 17 March 2011.

1. Mair, W. & Dillin, A. Aging and survival: the genetics of life span extension by dietary restriction. *Annu. Rev. Biochem.* **77**, 727–754 (2008).
2. Bishop, N. A. & Guarente, L. Genetic links between diet and lifespan: shared mechanisms from yeast to humans. *Nature Rev. Genet.* **8**, 835–844 (2007).
3. Honjoh, S., Yamamoto, T., Uno, M. & Nishida, E. Signalling through RHEB-1 mediates intermittent fasting-induced longevity in *C. elegans*. *Nature* **457**, 726–730 (2009).
4. Bishop, N. A. & Guarente, L. Two neurons mediate diet-restriction-induced longevity in *C. elegans*. *Nature* **447**, 545–549 (2007).
5. Carrano, A. C., Liu, Z., Dillin, A. & Hunter, T. A conserved ubiquitination pathway determines longevity in response to diet restriction. *Nature* **460**, 396–399 (2009).
6. Kapahi, P. *et al.* With TOR, less is more: a key role for the conserved nutrient-sensing TOR pathway in aging. *Cell Metab.* **11**, 453–465 (2010).
7. Panowski, S. H., Wolff, S., Aguilaniu, H., Durieux, J. & Dillin, A. PHA-4/Foxa mediates diet-restriction-induced longevity of *C. elegans*. *Nature* **447**, 550–555 (2007).
8. Di Marzo, V. & Matias, I. Endocannabinoid control of food intake and energy balance. *Nature Neurosci.* **8**, 585–589 (2005).
9. McPartland, J. M. & Glass, M. Functional mapping of cannabinoid receptor homologs in mammals, other vertebrates, and invertebrates. *Gene* **312**, 297–303 (2003).
10. Hardison, S., Weintraub, S. T. & Giuffrida, A. Quantification of endocannabinoids in rat biological samples by GC/MS: technical and theoretical considerations. *Prostaglandins Other Lipid Mediat.* **81**, 106–112 (2006).
11. Lehtonen, M., Reisner, K., Auriola, S., Wong, G. & Callaway, J. C. Mass-spectrometric identification of anandamide and 2-arachidonoylglycerol in nematodes. *Chem. Biodivers.* **5**, 2431–2441 (2008).
12. Di Marzo, V., Bifulco, M. & De Petrocellis, L. The endocannabinoid system and its therapeutic exploitation. *Nature Rev. Drug Discov.* **3**, 771–784 (2004).
13. McPartland, J. M., Matias, I., Di Marzo, V. & Glass, M. Evolutionary origins of the endocannabinoid system. *Gene* **370**, 64–74 (2006).
14. Leung, D., Saghatelian, A., Simon, G. M. & Cravatt, B. F. Inactivation of N-acyl phosphatidylethanolamine phospholipase D reveals multiple mechanisms for the biosynthesis of endocannabinoids. *Biochemistry* **45**, 4720–4726 (2006).
15. Fielenbach, N. & Antebi, A. *C. elegans* dauer formation and the molecular basis of plasticity. *Genes Dev.* **22**, 2149–2165 (2008).
16. Kirkham, T. C., Williams, C. M., Fezza, F. & Di Marzo, V. Endocannabinoid levels in rat limbic forebrain and hypothalamus in relation to fasting, feeding and satiation: stimulation of eating by 2-arachidonoyl glycerol. *Br. J. Pharmacol.* **136**, 550–557 (2002).
17. Izzo, A. A. *et al.* Basal and fasting/refeeding-regulated tissue levels of endogenous PPAR- α ligands in Zucker rats. *Obesity* **18**, 55–62 (2010).
18. Chen, D., Thomas, E. L. & Kapahi, P. HIF-1 modulates dietary restriction-mediated lifespan extension via IRE-1 in *Caenorhabditis elegans*. *PLoS Genet.* **5**, e1000486 (2009).
19. Kaeberlein, T. L. *et al.* Lifespan extension in *Caenorhabditis elegans* by complete removal of food. *Aging Cell* **5**, 487–494 (2006).
20. Banni, S. & Di Marzo, V. Effect of dietary fat on endocannabinoids and related mediators: consequences on energy homeostasis, inflammation and mood. *Mol. Nutr. Food Res.* **54**, 82–92 (2010).
21. Watts, J. L. & Browse, J. Genetic dissection of polyunsaturated fatty acid synthesis in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **99**, 5854–5859 (2002).
22. Valenti, M. *et al.* The endocannabinoid system in the brain of *Carassius auratus* and its possible role in the control of food intake. *J. Neurochem.* **95**, 662–672 (2005).
23. Soderstrom, K., Tian, Q., Valenti, M. & Di Marzo, V. Endocannabinoids link feeding state and auditory perception-related gene expression. *J. Neurosci.* **24**, 10013–10021 (2004).
24. Breunig, E. *et al.* The endocannabinoid 2-arachidonoyl-glycerol controls odor sensitivity in larvae of *Xenopus laevis*. *J. Neurosci.* **30**, 8965–8973 (2010).
25. De Petrocellis, L., Melck, D., Bisogno, T., Milone, A. & Di Marzo, V. Finding of the endocannabinoid signalling system in *Hydra*, a very primitive organism: possible role in the feeding response. *Neuroscience* **92**, 377–387 (1999).
26. Elphick, M. R. & Egertova, M. The phylogenetic distribution and evolutionary origins of endocannabinoid signalling. *Handb. Exp. Pharmacol.* **168**, 283–297 (2005).
27. Elphick, M. R. BfCBP: a cannabinoid receptor ortholog in the cephalochordate *Branchiostoma floridae* (Amphioxus). *Gene* **399**, 65–71 (2007).
28. Sulston, J., Hodgkin, J. & Wood, W. B. in *The Nematode Caenorhabditis elegans* 587–606 (Cold Spring Harbor Laboratory, 1988).
29. Held, J. M. *et al.* DAF-12-dependent rescue of dauer formation in *Caenorhabditis elegans* by (25S)-cholestenoic acid. *Aging Cell* **5**, 283–291 (2006).
30. Lithgow, G. J., White, T. M., Melov, S. & Johnson, T. E. Thermotolerance and extended life-span conferred by single-gene mutations and induced by thermal stress. *Proc. Natl Acad. Sci. USA* **92**, 7540–7544 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Some nematode strains used in this work were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources (NCRR). We would like to thank N. J. Harrison, A. Olsen and P. Kapahi. M.L. was supported by NIH training grant T32 AG000266 and NIH grant R01 AG029631. GC-MS analysis was made possible through the Mass Spectrometry and Imaging Technologies Core supported by NIH grant PL1-AG032118. This work was supported by a Larry L. Hillblom Foundation grant and NIH grants to G.J.L. (UL1 DE019608, supporting the Interdisciplinary Research Consortium on Geroscience and R01 AG029631) and by NIH grants R21 AG030192 and R01 AG036992 to M.S.G.

Author Contributions M.L., J.M.H., B.W.G., G.J.L. and M.S.G. conceived of and planned experiments. M.L., M.C.V., I.M.K., J.B.G. and M.S.G. performed experiments. M.L. and M.S.G. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.S.G. (mgill@scripps.edu).

METHODS

Chemicals. Palmitoyl ethanolamide (PEA), oleoyl ethanolamide (OEA), linoleoyl ethanolamide (LOEA), arachidonoyl ethanolamide (AEA), PEA-*d*₂, OEA-*d*₄ and AEA-*d*₄ were obtained from Cayman Chemical. Eicosapentaenoyl ethanolamide (EPEA) and palmitoleoyl ethanolamide (POEA) were obtained from Enzo Life Sciences. BSTFA was from Sigma Aldrich. All solvents were of GC-MS grade and all other reagents and solvents were of the highest grade available.

Nematode strains. The following nematode strains used in this work were provided by the *Caenorhabditis* Genetics Center, University of Minnesota, except *age-1(m875) II* which was obtained from D. Chen: Bristol N2 (wild type), *daf-2(e1368) III*, *daf-2(m41) III*, *daf-2(e1370) III*, *daf-2(e1371) III*, *age-1(hx546) II*, *daf-7(e1372) III*, *daf-4(e1364) III*, *unc-31(e928) IV*, *unc-64(e246) III*, *daf-16(mu86) I*, *pdk-1(sa709) X*, *akt-1(ok525) V*, *daf-8(e1393) I*, *daf-11(m47) V*, *daf-28(sa191) V*, *eri-1(mg366) IV*; *lin-15B(n744) X*, *fat-4(wa14) IV*, *rsk-1(ok1255) III*; *C02H7.2(ok2068) X*, *nhr-49(gk405) I*, *ocr-2(ok1711) IV*, *osm-9(ok1677) IV*, *ser-1(ok345) X*, *ser-4(ok512) III*.

Nematode culture. For routine culture, worms were maintained using standard culture methods²⁸. For mass cultures, 10,000 worms were grown on 10-cm NGM plates seeded with 2 ml concentrated *E. coli* OP50 and starved L1s from eggs prepared by sodium hypochlorite treatment³¹.

Generation of transgenic lines. Promoter-GFP fusions were generated essentially as described previously³². To generate the *Pnape-1::GFP* construct, approximately 3 kb of the 5' region of the *nape-1* gene up to the seventh codon after the translational start (primer f1 5'-TCACGTGAGAAATAGTCGCTGG-3' to primer r1 5'-TGAGGAGTTGTCGGTAGGCTCAT-3') was fused to a *GFP::unc-54* 3' UTR fragment of pPD95.75 (gift from A. Fire). To generate the *Pfaah-1::GFP* construct, approximately 3 kb of the 5' region of the *faah-1* gene up to and including the translational start (primer f1 5'-CCATGTAGAGAGCCTTCCGACAT-3' to primer r1 5'-CATGATGACCTTGAAATACTGAAAATTGAA-3') was fused to a *GFP::unc-54* 3' UTR sequence.

To generate the *Prps-5::faah-1* overexpressor lines, a *faah-1* genomic fragment, encompassing the entire open reading frame from the start codon to the end of the annotated 3' UTR (primer f1 5'-ATGATTTTTTACTTGGTGCTTCTCGTTT TG-3' to primer r2 5'-GGACAAGGAATGGTGGACTTCGG-3'), was PCR fused³³ to the ubiquitous constitutive promoter *Prps-5* (ref. 34) (primer f1 5'-CG AAACGGTAACGTGAAGAATTCAATA-3' to primer r1 5'-CTTGCAAAAT AACAACCTCAGTATAGT-3'). The resulting 6-kb PCR fragment was micro-injected into the gonad of young adult worms and resulting transgenic animals were identified in the next generation by their expression of a co-injection marker. At least five independent transgenic lines were analysed. A representative line was chosen for integration using trimethylpsoralen/UV mutagenesis. Two independent integrants were isolated and named *rfls22[Prps-5::faah-1, Podr-1::dsRED, Punc-25::mRFP]X* and *rfls23[Prps-5::faah-1, Podr-1::dsRED, Punc-25::mRFP]X*, respectively. These lines were outcrossed to wild type six times before analysis.

To generate worm lines with tissue-specific expression of *faah-1*, we used the *faah-1* genomic fragment described above and discretely PCR spliced it to promoter fragments that drive expression specifically in pharyngeal muscle (*Pmyo-2*), body wall muscle (*Pmyo-3*) and neurons (*Punc-119*). Primers for *Pmyo-2*: f1, 5'-GGTGGTGACAGTAACGTCTGT-3'; r1, 5'-CATTTCTGTGTCTGACG ATCGAGG-3'. Primers for *Pmyo-3*: f1, 5'-CACTTCCGGCGCCCTGAATCT AA-3'; r1, 5'-CATTTCTAGATGGATCTAGTGGTGGTGGG-3'. Primers for *Punc-119*: f1, 5'-GGCTCCAATCGGAACGCGAACA-3'; r1, 5'-CATATATGC TGTGTAGCTGAAAATTTGGGGATTATGGG-3'.

Mass culture for GC-MS of larval stages. For larval cultures of N2 and *daf-2(e1368)* for GC-MS analysis, the NGM plates included 5 µg ml⁻¹ lathosterol to facilitate reproductive growth and prevent dauer formation of *daf-2(e1368)* when grown in mass culture. To generate larval cultures, N2 and *daf-2(e1368)* were grown at 15 °C on NGM plates to generate 1 × 10⁷ arrested L1s which were then grown in liquid culture at 25 °C until the appropriate larval stage. Supplementary Fig. 17 shows that NAE levels are not different when worms are grown on solid media and liquid culture, indicating that liquid culture conditions supplemented with the appropriate amount of food do not induce dietary restriction. Arrested L1s were inoculated into 2-l glass flasks (no more than 1 × 10⁶ worms per flask) containing 90 ml S-medium (plus cholesterol, final concentration 5 µg ml⁻¹) and 10 ml of concentrated *E. coli* OP50. The liquid cultures were then incubated in a shaking incubator at 25 °C, 150 r.p.m. The number of animals at each larval stage required to yield 0.7 g wet weight was calculated using data from ref. 35 and flasks were inoculated accordingly (in the order larval stage, time of harvest, number of worms: N2: L1, 6 h, 5 × 10⁶ worms; L2, 15 h, 2.5 × 10⁶ worms; L3, 21 h, 1 × 10⁶ worms; L4, 28 h, 5 × 10⁵ worms; young adult, 36 h, 2.5 × 10⁵ worms; gravid adult, 48 h, 2.5 × 10⁵ worms; *daf-2(e1368)*: L1, 6 h, 5 × 10⁶ worms; L2, 15 h, 2.5 × 10⁶ worms; L2d, 25 h, 2.5 × 10⁶ worms; dauer, 48 h, 1 × 10⁶ worms). Worms were

harvested by low-speed centrifugation and washed three times with S-basal before the worm pellet was snap frozen in liquid nitrogen.

To compare NAE levels in fed and starved L1s, N2 animals were grown on 10-cm NGM plates. Eggs harvested by sodium hypochlorite treatment were allowed to hatch in S-basal in the absence of food and maintained at 20 °C for 24 h. At this point arrested L1s were inoculated into 2-l glass flasks (no more than 1 × 10⁶ worms per flask) containing 100 ml S-medium (plus cholesterol, final concentration 5 µg ml⁻¹) or 90 ml S-medium plus 10 ml of concentrated *E. coli* OP50 for starved and fed cultures, respectively. The liquid cultures were then incubated in a shaking incubator at 20 °C, 150 r.p.m. for 6 h after which worms were harvested by low-speed centrifugation and washed three times with S-basal before the worm pellet was snap frozen in liquid nitrogen.

Mass cultures with FAAH inhibitor URB597. URB597 (Cayman Chemical) was re-suspended in DMSO to a final concentration of 20 mM and further diluted in S-basal for spotting onto plates. Ten-centimetre NGM plus *E. coli* plates were inoculated with 10,000 arrested L1s and grown at 20 °C for 48 h. At this point animals were harvested and transferred to NGM plus *E. coli* plates containing either DMSO vehicle or 10 µM URB597 and incubated for a further 24 h at 20 °C before harvesting.

Mass culture for dietary restriction and re-feeding in adult animals. To generate mass cultures under dietary restriction conditions we used a modification of the protocol described previously³⁶. Ten-centimetre NGM plates were inoculated with 10,000 arrested L1s and grown at 20 °C for 48 h. At this point worms were washed off and transferred to plates containing 10 µg ml⁻¹ 2'-deoxy-5-fluorouridine (FUDR) to inhibit progeny production. After a further 24 h at 20 °C, adult worms were harvested and inoculated into 2-l flasks (75,000 worms per flask) containing S-medium (plus 5 mg ml⁻¹ cholesterol, 100 µg ml⁻¹ FUDR, 50 µg ml⁻¹ carbenicillin). Flasks were supplemented with concentrated OP50 to give an optical density at 600 nm (OD₆₀₀) of 2.5 for fed conditions and OD₆₀₀ = 0.2 for dietary restriction conditions. The liquid cultures were then incubated in a shaking incubator at 20 °C, 150 r.p.m. for either 12 or 24 h after which worms were harvested by low-speed centrifugation and washed three times with S-basal before the worm pellet was snap frozen in liquid nitrogen. For re-feeding experiments, worms were incubated for 12 h under dietary restriction conditions and then harvested and re-inoculated into flasks containing *E. coli* OP50 at OD₆₀₀ = 2.5 for an additional 12 h before harvesting.

Lipid extraction. Lipid extracts were generated by a modification of the method of ref. 37. A total of 600 mg frozen worm pellets were thawed on ice in 4 ml methanol in 50-ml glass centrifuge tubes and then subjected to 4 × 1 min sonication on ice. After sonication, 30 ng of the internal standards PEA-*d*₂, OEA-*d*₄ and AEA-*d*₄ were added followed by 8 ml chloroform and 4 ml 0.5 M KCl/0.08 M H₃PO₄ to a final ratio of 1:2:1. Samples were vortexed and then sonicated in an ultrasonic water bath for 15 min. After vortexing for 2 × 1 min, samples were centrifuged for 10 min at 2,000g to separate the phases. The lower phase was collected into a clean glass tube, dried under nitrogen and re-suspended in 4 ml hexane.

Solid-phase fractionation. The total lipid extracts were fractionated on Sep-Pak Classic silica columns (Waters Corporation) to collect the fatty acid amide and monoglyceride fraction using a modification of the method of ref. 37. Sequential elutions were performed using 2 ml 99:1 hexane:acetic acid v/v, 15 ml 90:10 hexane:ethyl acetate v/v, 10 ml 80:20 hexane:ethyl acetate v/v, 5 ml 70:30 hexane:ethyl acetate v/v with the final elution of 5 ml 2:1 chloroform:isopropanol v/v containing the fatty acid amide/monoglyceride fraction. This sample was dried under nitrogen and re-suspended in 1.5 ml hexane. The volumes for the solvent cuts were optimized by analysing each fraction by thin-layer chromatography.

Gas chromatography mass spectrometry. Samples were analysed by GC-MS using GC conditions based on those described previously¹⁰. Samples were analysed in duplicate, corresponding to at least 200 mg worms per run, after derivatization with BSTFA for 1 h. Just before GC-MS analysis, samples were dried under N₂ and re-suspended in 1 µl hexane for injection. GC-MS analysis was performed using a Varian 2100T ion-trap GC/MS/MS with a 3900 GC (Varian Inc.) operating in splitless mode with a VF-5ms capillary column (30 m × 0.25 mm internal diameter, 5% phenyl-95% methyl polysiloxane, 0.25 µm film thickness; Varian, Inc.). GC conditions: injector 250 °C. Initial column temperature was 150 °C for 1 min and then ramped at 20 °C per min to 300 °C and held for 15 min. MS conditions: analytes were chemically ionized using acetonitrile vapour. Data was collected using the MRM mode in Varian System Control software v6.40. Precursor ions were isolated using an isolation window of 3. MS/MS fragmentation was performed in the ion trap using an excitation storage level of 164.0, excitation amplitude of 80, and non-resonant collision energy for all analytes. Three analysis segments were performed in each run with the endogenous NAEs and labelled internal standard with equivalent carbon atoms analysed during each segment. The C16 NAEs (POEA, PEA and PEA-*d*₄) were typically analysed for the first 7.5 min of each run, the C18 NAEs (LOEA, OEA and OEA-*d*₂)

segment was from 7.5 to 8.0 min, and the C20 NAEs (EPEA, AEA and AEA- d_4) were analysed for the rest of the run. The MS scans cycled between each analyte continuously with a total cycle time of 0.72 s for three analytes per segment.

Peak areas were generated by manually integrating the extracted ion chromatogram for the MS/MS fragment ion of interest (POEA m/z 370, 280; PEA m/z 372, 282; PEA- d_4 m/z 376, 286; LOEA m/z 396, 306; OEA m/z 398, 308; OEA- d_2 m/z 400, 310; EPEA m/z 418, 328; AEA m/z 420, 330; AEA- d_4 m/z 424, 334) using Varian MS Data Review v6.4. The fragment ion chosen for each transition corresponds to the loss of O-TMS ($[M + H - 90]^+$), which was the base peak in all MS/MS spectra and had the maximum signal to noise and dynamic range.

Dauer assays. Dauer assays were performed as previously described²⁹. NAEs were re-suspended in ethanol to a final concentration of 10 mM. For dose-range experiments serial dilutions were made to yield 10, 5, 2, 1 and 0.5 mM. 15 μ l of each working solution of NAEs were added to 135 μ l S-basal before being spotted onto a 3-ml NGM. Equal distribution of the NAE throughout the agar was assumed to yield final concentrations of 50, 25, 10, 5 and 2.5 μ M. Plates were then scored for the number of animals at each developmental stage (daughters, L3, L4, young adults and gravid adults).

Thermotolerance assays. Thermotolerance assays were performed as previously described³⁰. For experiments with the *faah-1* overexpression lines, first day adults were transferred from the growth temperature (20 °C) to a 35 °C heat shock incubator. For experiments with EPEA, worms were transferred to plates containing 50 μ M EPEA at the L4 stage and after 24 h were shifted to the 35 °C incubator.

Lifespan analysis. Lifespan analysis was performed as previously described³⁰. On the first day of adulthood animals were transferred to plates containing 10 μ g ml⁻¹ FUDR, to inhibit progeny production, with each condition assayed in duplicate with 50 worms per plate. Worms were transferred to new plates every 2 days and transferred to plates without FUDR at day 8. Animals that crawled off the plate, experienced internal hatching or exhibited vulval protrusion were scored as censored data. Survival analysis was performed using Prism 4 software (Graphpad Software); Kaplan–Meier survival curves were plotted for each lifespan assay and compared using the log-rank test. For dietary restriction experiments we used the method of ref. 18, with animals transferred to dietary restriction plates as first-day adults. To examine the effect of EPEA on lifespan, worms from a synchronous lay were transferred to plates containing 50 μ M EPEA at the L4 stage. After 24 h worms were transferred to plates containing FUDR. For the lifespan analysis worms were transferred to fresh EPEA every 2 days. For dietary restriction lifespans in the presence of EPEA, worms from a synchronous lay were transferred to EPEA plates at the L4 stage and moved to the dietary restriction plates plus 10 μ g ml⁻¹ FUDR as first-day adults. Worms were transferred every day for the first 8 days to prevent food depletion in the dietary restriction condition and every 2 days thereafter. For *rsk-1* lifespans with EPEA, worms were moved to EPEA plates at L4 and moved

to FUDR plates 24 h later. Worms were transferred to fresh EPEA plates every day for the first 8 days and every 2 days thereafter.

RNA interference. RNA interference by feeding and by soaking was performed as previously described^{38,39}. For mass culture RNAi by soaking, approximately 2,500 L4 worms were incubated for 24 h in 1 ml of M9 buffer with 0.5 μ g μ l⁻¹ of dsRNA synthesized *in vitro* with the megascript kit (Ambion Inc.). Soaking was performed in 50-ml glass tubes with slight shaking. The resulting P0 young adults were dispersed onto mass culture plates and allowed to produce F₁ offspring. The mass cultures were harvested for analysis before the F₁ produced eggs. The *faah-1* (B0218.1), *nape-1* (Y37ARE11.4) and *pha-4* (F38A6.1) RNAi clones were from the *C. elegans* ORF-RNAi Library (Open Biosystems). The sequences of the open-reading frames for these genes are available at <http://www.wormbase.org>.

Length measurement. Worms were grown on either control RNAi or *faah-1* RNAi for two generations. Eggs from a 15-min synchronous lay were collected and incubated at 20 °C. Each day, 20 larvae were mounted on a glass slide and anaesthetized with 0.1% sodium azide. Bright-field images were captured on an Olympus IX70 inverted microscope using an Olympus DP70 digital camera. Body size was measured using Image J software (NIH) and calibrated against a stage micrometer.

Fertility assays. Synchronous populations of worms were cultured at 20 °C until the L4 stage. Individual worms were then transferred and maintained each on a separate plate and transferred daily. Progeny were left to develop for 24 h before counting. For fertility in the presence of EPEA, eggs were transferred to control or 50 μ M EPEA plates and after 48 h single worms were transferred daily to individual plates containing vehicle or EPEA.

31. Fabian, T. J. & Johnson, T. E. Production of age-synchronous mass cultures of *Caenorhabditis elegans*. *J. Gerontol.* **49**, B145–B156 (1994).
32. Hobert, O. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. *Biotechniques* **32**, 728–730 (2002).
33. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. & Pease, L. R. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**, 61–68 (1989).
34. Dupuy, D. *et al.* Genome-scale analysis of *in vivo* spatiotemporal promoter activity in *Caenorhabditis elegans*. *Nature Biotechnol.* **25**, 663–668 (2007).
35. Knight, C. G., Patel, M. N., Azevedo, R. B. & Leroi, A. M. A novel mode of ecdysozoan growth in *Caenorhabditis elegans*. *Evol. Dev.* **4**, 16–27 (2002).
36. Mair, W., Panowski, S. H., Shaw, R. J. & Dillin, A. Optimizing dietary restriction for genetic epistasis analysis and gene discovery in *C. elegans*. *PLoS ONE* **4**, e4535 (2009).
37. Sultana, T. & Johnson, M. E. Sample preparation and gas chromatography of primary fatty acid amides. *J. Chromatogr. A* **1101**, 278–285 (2006).
38. Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
39. Timmons, L., Tabara, H., Mello, C. C. & Fire, A. Z. Inducible systemic RNA silencing in *Caenorhabditis elegans*. *Mol. Biol. Cell* **14**, 2972–2983 (2003).

A novel tumour-suppressor function for the Notch pathway in myeloid leukaemia

Apostolos Klinakis^{1*}, Camille Lobry^{2*}, Omar Abdel-Wahab³, Philmo Oh², Hiroshi Haeno⁴, Silvia Buonamici^{2,8}, Inge van De Walle⁵, Severine Cathelin², Thomas Trimarchi², Elisa Araldi², Cynthia Liu², Sherif Ibrahim², Miroslav Beran⁶, Jiri Zavadil⁷, Argiris Efstratiadis¹, Tom Taghon⁵, Franziska Michor⁴, Ross L. Levine³ & Iannis Aifantis²

Notch signalling is a central regulator of differentiation in a variety of organisms and tissue types¹. Its activity is controlled by the multi-subunit γ -secretase (γ SE) complex². Although Notch signalling can play both oncogenic and tumour-suppressor roles in solid tumours, in the haematopoietic system it is exclusively oncogenic, notably in T-cell acute lymphoblastic leukaemia, a disease characterized by Notch1-activating mutations³. Here we identify novel somatic-inactivating Notch pathway mutations in a fraction of patients with chronic myelomonocytic leukaemia (CMML). Inactivation of Notch signalling in mouse haematopoietic stem cells (HSCs) results in an aberrant accumulation of granulocyte/monocyte progenitors (GMPs), extramedullary haematopoiesis and the induction of CMML-like disease. Transcriptome analysis revealed that Notch signalling regulates an extensive myelomonocytic-specific gene signature, through the direct suppression of gene transcription by the Notch target *Hes1*. Our studies identify a novel role for Notch signalling during early haematopoietic stem cell differentiation and suggest that the Notch pathway can play both tumour-promoting and -suppressive roles within the same tissue.

Notch is essential for the emergence of definitive haematopoiesis⁴, controls HSC differentiation to the T-cell lineage^{5,6} and is a major oncogene, as most patients with T-cell acute lymphoblastic leukaemia harbour activating *Notch1* mutations⁷. To study haematopoiesis in the absence of any Notch-derived signal (as mammals express four different Notch receptors), we targeted Nicastrin (*Ncstn*), a member of the γ SE complex and one of the few non-redundant members of the pathway (Supplementary Fig. 1). We crossed the *Ncstn*^{flf} mice to both an inducible (Mx1-cre)⁸ and a haematopoietic-specific (Vav-cre)⁹ recombinase strain. Both modes of deletion (referred to herein as *Ncstn*^{-/-}) produced identical phenotypes. Unexpectedly, none of the *Ncstn*^{-/-} mice survived longer than 20 weeks. Further analysis revealed a striking peripheral blood leukocytosis and monocytosis with enlargement of the spleen (Fig. 1a, b and Supplementary Figs 2 and 3). Histological analysis of the spleen showed a marked expansion of the red pulp with diffuse infiltration by myeloid and monocytic cells. The infiltrating myeloid cells were partly myeloperoxidase positive and CD11b⁺ and/or Gr1⁺ (Fig. 1b and Supplementary Figs 2 and 3). The increase in monocyte numbers was also observed in the bone marrow and liver (Fig. 1a and data not shown). Taken together, these findings were diagnostic for a myeloproliferative/myelodysplastic process and reminiscent of human CMML. CMML is a myeloid malignancy, classified as a myeloproliferative disorder/myelodysplastic syndrome overlap, which is characterized by monocytosis, myeloproliferation, variable bone marrow dysplasia and a high rate of progression to acute myeloid leukaemia¹⁰.

Because monocytes and granulocytes originate from the GMP subset, we examined the stem and progenitor cell populations in the bone marrow. *Ncstn* deletion lead to an enlargement of the Lineage^{neg}Sca1⁺c-Kit⁺ (LSK), specifically the LSK CD150⁺CD48⁺ subset, a population shown to have a myeloid commitment bias¹¹. This differentiation bias was coupled to a significant reduction of the lymphoid-biased multipotential progenitor population (LMPP)¹² (Supplementary Fig. 4). Moreover, there was a striking increase in the absolute numbers of both bone marrow and spleen GMP cells (Fig. 1d) coupled to a decrease of the megakaryocyte-erythrocyte progenitor population. This apparent predisposition towards GMP-derived lineages was also evident *in vitro*, as *Ncstn*^{-/-} progenitors generated more granulocyte-macrophage and macrophage colonies (Supplementary Fig. 5). Further studies also revealed a striking ability of the *Ncstn*^{-/-} progenitors to replate serially (Supplementary Fig. 6), suggesting an increase in their self-renewal potential. Consistent with this idea, whole transcriptome profiling of *Ncstn*^{-/-} GMP progenitors revealed enrichment of a 'leukemic self-renewal' signature¹³. Finally, bone marrow transplantation assays demonstrated that the effects of *Ncstn* deletion were cell autonomous (Supplementary Fig. 7)¹⁴.

Although the γ SE complex has other substrates², we focused on Notch signalling because of our finding that *Ncstn* deletion led to known Notch^{-/-} phenotypes, including a block in T-cell differentiation (Supplementary Fig. 8)¹⁵. To prove a connection to Notch signalling, we generated animals that conditionally lack the expression of three out of four Notch receptors (*Mx1-cre*⁺*N1^{flf}N2^{flf}N3^{-/-}*)¹⁶. Strikingly, triple *Notch1/2/3* deletion copied the *Ncstn*^{-/-} phenotypes (Supplementary Fig. 9). *Notch1-3^{-/-}* mice developed both CMML-like symptoms, and significant enlargement of the GMP population. *Notch3* expression was dispensable, as simultaneous deletion of only *Notch1* and *Notch2* led to an identical CMML-like pathology. However, introduction of a single wild-type (WT) *Notch1* or *Notch2* allele was able to suppress the disease phenotype. Consistent with the importance of Notch1–2 receptor signalling in these stages, quantitative PCR studies revealed expression of *Notch1* and *Notch2*, but not *Notch3*, in WT stem and progenitor cells (Supplementary Fig. 10).

We next sought to delineate the mechanism by which Notch directs the regulation of early haematopoiesis. Deletion of *Ncstn* did not lead to any alterations in the GMP cell-cycle status or cell death rate (not shown). We hypothesized that γ SE complex/Notch signals actively suppress a GMP-specific gene expression program. We sorted LSK and GMP cells and studied their transcriptome. This analysis revealed a statistically significant de-repression of an extended myeloid-specific program¹² in *Ncstn*^{-/-} LSK cells (Fig. 2a and Supplementary Fig. 11).

¹Biomedical Research Foundation, Academy of Athens, Athens, Greece. ²Howard Hughes Medical Institute and Department of Pathology, New York University School of Medicine, New York, New York 10016, USA. ³Human Oncology and Pathogenesis Program and Leukemia Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York 10016, USA. ⁴Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, and Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ⁵Department of Clinical Chemistry, Microbiology and Immunology, Ghent University Hospital, Ghent University, Ghent, Belgium. ⁶Department of Leukemia, M.D. Anderson Cancer Center, Houston, Texas 77030, USA. ⁷Department of Pathology, NYU Cancer Institute and Center for Health Informatics and Bioinformatics, NYU Langone Medical Center, New York, New York 10016, USA. ⁸Novartis Institutes for Biomedical Research, Cambridge, Massachusetts 02139, USA.

*These authors contributed equally to this work.

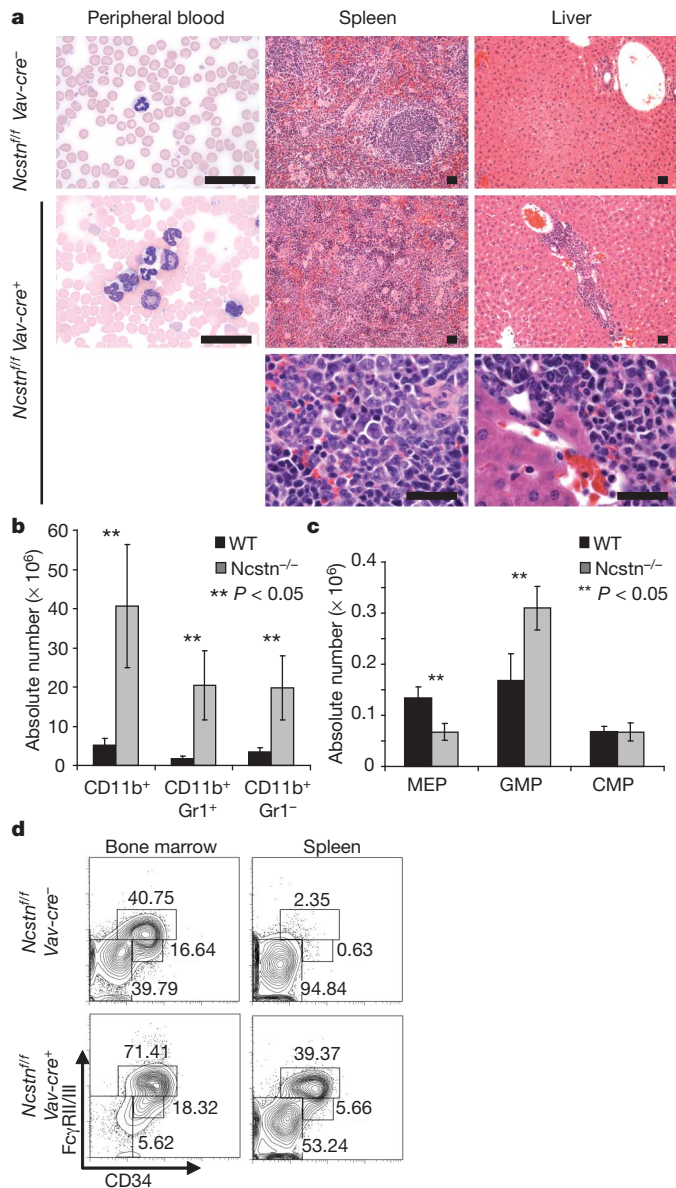


Figure 1 | *Ncstn* deficiency leads to CMML-like disease and a significant enlargement of the GMP progenitor population. **a**, Histological analysis showing accumulation of monocytes and granulocytes in peripheral blood (Wright–Giemsa staining), spleen and liver (haematoxylin and eosin staining). A magnification of each infiltrant is shown in the lowest panels. **b**, Absolute numbers of each monocytic/granulocytic subset from the spleen of control and *Ncstn^{fl/fl} Vav-cre⁺* littermate animals (12 weeks of age, mean \pm s.d., $n = 10$). **c**, Absolute numbers of each progenitor subpopulation in the bone marrow (mean \pm s.d., $n = 10$). **d**, Detailed FACS analysis of bone marrow and spleen myeloid progenitor (myeloid progenitors: Lin⁻/c-Kit⁺/Sca-1⁻) populations of *Ncstn^{fl/fl} Vav-cre⁺* and *Ncstn^{fl/fl} Vav-cre⁻* littermates showing a significant enlargement of the GMP (FcγRII/III⁺, CD34⁺) subset.

Gene-set enrichment analysis demonstrated a significant enrichment of myeloid-specific gene-sets within the *Ncstn^{-/-}* LSK gene signature (Fig. 2b and Supplementary Table 1). Further dissection of the LSK subset showed that the GMP gene expression program was initiated as early as the CD150⁺ HSC stage of differentiation, and persisted at the CD150⁻ subset, which included MPPs (Supplementary Fig. 11). Furthermore, we were able to show that gene expression in LSKs purified from an inducible Notch1 gain-of-function genetic model (*Eflα1-lsl-Notch1^{IC}Mx1-cre⁺*)¹⁷ was inversely correlated with the loss-of-function signature seen in *Ncstn^{-/-}* LSK progenitors (Fig. 2a). Notch1^{IC} expression led to suppression of myeloid-specific genes, suggesting that activation of Notch

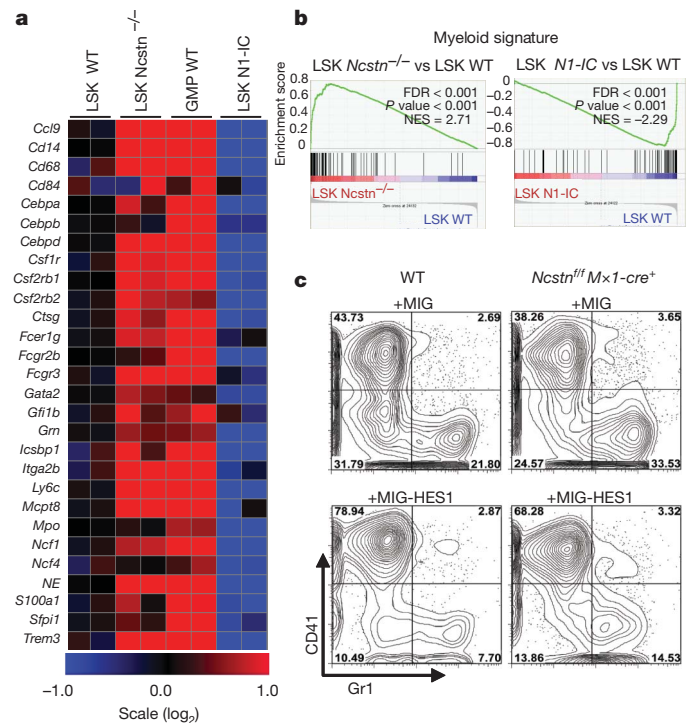


Figure 2 | Notch signalling suppresses an extensive myeloid gene expression program through the induction of the transcriptional repressor Hes1.

a, Heat map showing regulation of genes representative of the myeloid signature from the indicated cell populations and mice. **b**, Expression data were analysed for lists of genes positively involved in myelopoiesis using gene-set enrichment analysis. Enrichment plots show upregulation of myeloid-specific genes in *Ncstn^{fl/fl}Mx1-cre⁺* and downregulation of myeloid-specific genes in *Notch1^{IC} Mx1-cre⁺* LSK cells (compared with WT counterparts). **c**, Purified cKit⁺ progenitors from WT and *Ncstn^{fl/fl}Mx1-cre⁺* mice were transduced with retroviruses encoding Hes-1 or empty vector, subsequently plated on methylcellulose for 7 days and analysed for expression of myeloid or megakaryocyte differentiation markers (Gr1, CD41). A representative of four experiments is shown.

signalling can alter the transcription and differentiation of uncommitted HSC and MPP cells. Gene-set enrichment analysis further supported these findings as it showed a negative correlation between the Notch1^{IC} LSKs and previously reported myeloid gene expression signatures (Fig. 2a, b, Supplementary Table 2 and Supplementary Fig. 11).

As Notch is thought to function primarily as a transcriptional activator, we hypothesized that its suppressive effects on GMP-specific gene expression could be explained by the induction of a transcriptional repressor. A search for such a molecule revealed *Hes1*, a known Notch target and a transcriptional repressor (Supplementary Fig. 11a). Interestingly, *Hes1* was previously suggested to play a role in myeloid leukaemia, as a downstream effector of the *Junb* tumour suppressor in an animal model of chronic myeloid leukemia¹⁸. To test the potential involvement of *Hes1* directly, we used *in vitro* differentiation assays and showed that *Hes1* ectopic expression was sufficient to direct differentiation away from the granulocyte/monocyte lineage (Fig. 2c). Furthermore, *Hes1* expression suppressed the expression of key granulocyte/monocyte commitment genes such as *Cebpa* and *Pu.1* (Supplementary Fig. 12a). In agreement with this finding, we identified putative *Hes* (N-box) binding sites on the promoters of both genes. Reporter and chromatin immunoprecipitation assays proved direct binding of *Hes1* on these promoters and suppression of transcription (Supplementary Fig. 12b–e).

These findings suggested that Notch (or *Hes1*) hyper-activation could suppress CMML-like disease developing in *Ncstn^{-/-}* animals. To test this, we used the previously described *Eflα1-lsl-Notch1^{IC}* mice¹⁷. We generated *Notch1^{IC}Ncstn^{fl/fl}Mx1-Cre⁺* animals and analysed both GMP accumulation and disease progression. Notch1^{IC} expression was

sufficient to suppress both GMP expansion and disease development significantly (Fig. 3a, b and Supplementary Fig. 13). Interestingly, Notch1^{IC} expression drove progenitor commitment towards the lymphoid (T cell) and megakaryocyte–erythrocyte progenitor lineages¹⁹. However, Notch1^{IC} expression did not affect cell-cycle kinetics within the GMP subset (Supplementary Fig. 13). To uncouple differentiation to the GMP subset from effects on GMP homeostasis, we purified WT GMP and plated them on stroma in the presence or absence of Notch ligands (Dll1 and Dll4). A striking increase of apoptosis rate was noted in the presence of Dll1–4 (Fig. 3c). These experiments suggest that Notch ectopic expression can affect both the commitment to the GMP subset and the survival of already committed GMP progenitors.

Our studies so far demonstrate that Notch controls murine myelopoiesis and that its deletion leads to GMP expansion and monocytic disease. To prove that Notch is also important in human haematopoiesis and leukemogenesis, we initially cultured purified human CD34⁺CD38[−]Lin[−] bone marrow and cord blood stem and progenitor cells on stroma expressing different Jagged and Delta-like Notch ligands²⁰. We found that expression of Notch ligands efficiently suppressed differentiation of human multipotential progenitors towards both the granulocyte (CD15⁺) and monocyte (CD14⁺) lineages (Supplementary Figs 14 and 15). This suppression was prevented when the activity of the γ SE complex was suppressed using either small molecule inhibitors or by the expression of a dominant negative *MAML1* mutant.

To gain further insights into the role of Notch in human CMML, we extensively sequenced many γ SE/Notch pathway genes. Exon resequencing of specimens from patients with CMML (Supplementary Table 3) identified a substantial fraction (six novel mutations in 5 out of 42 patients) harbouring somatic heterozygous mutations in

multiple Notch pathway genes including *NCSTN*, *APH1*, *MAML1* and *NOTCH2* (Fig. 4 and Supplementary Table 4). In addition, several other putative mutations (single nucleotide variants) were detected, which are not annotated as known germline single nucleotide polymorphisms but for which we could not prove a somatic origin (Supplementary Table 5). The validated somatic mutations were only observed in CMML, as re-sequencing of 47 samples from patients with myeloproliferative disorders (polycythemia vera or myelofibrosis) did not reveal somatic mutations in the Notch pathway (Fig. 4b). Importantly, CMML specimens with Notch mutations also had somatic alterations in well-characterized myeloid oncogenic lesions²¹,

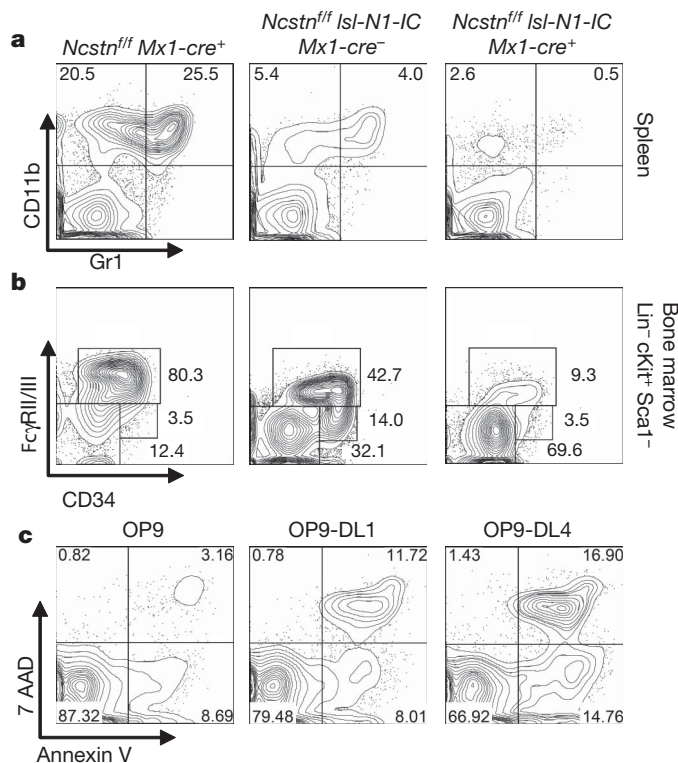


Figure 3 | Ectopic expression of Notch1-IC is able to prevent CMML-like disease in *Ncstn*^{−/−} mice. **a**, PolyI:polyC-induced Notch1-IC expression in *Ncstn*^{−/−}Isl-N1-IC *Mx1-Cre*⁺ animals suppresses myeloid cells in spleen. **b**, Notch1-IC expression suppresses GMP progenitor population in bone marrow. **c**, Induction of cell death in WT GMP cells cultured in the presence (OP9-DL1–4) or absence (OP9) of Notch ligands. Cell death was measured by the combination of 7AAD and annexin V staining 48 h after co-culture initiation. For **a–c**, a representative of more than three experiments is shown.

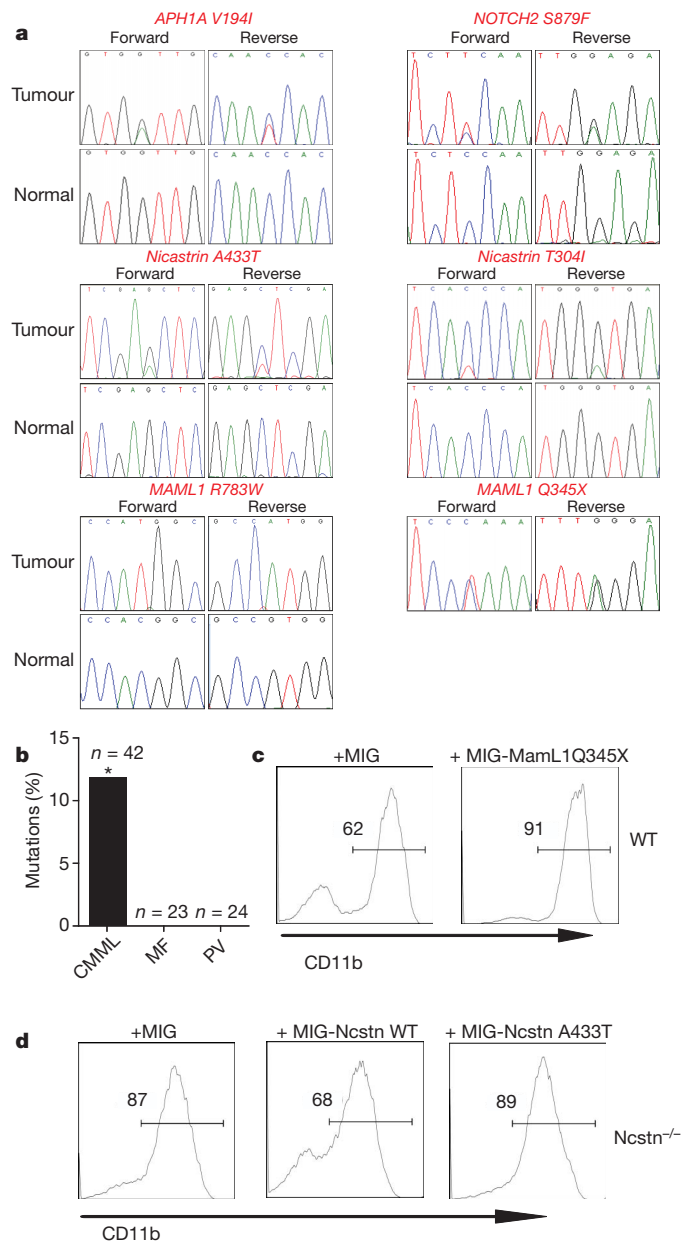


Figure 4 | Novel, loss-of-function Notch pathway mutations in human CMML. **a**, Sequence traces of identified Notch pathway mutations in tumours from patients with CMML but not in normal tissues show somatic origin. **b**, Comparison of the percentage of Notch pathway mutations in specimens from patients with CMML, myelofibrosis and polycythemia vera. The asterisk denotes that only verified somatic CMML mutations are included. **c**, OP9-DL1 co-culture of WT LSK cells infected with specified constructs. Analysis of the CD11b⁺ population was studied 14 days after the initiation of the culture. **d**, A similar experiment as in **c**, using LSK *Ncstn*^{−/−} progenitors infected with the specified constructs. In all cases, a representative of more than three experiments is shown.

including *JAK2*, *KRAS*, *TET2* and *ASXL1*, suggesting mutational cooperation between Notch signalling and other oncogenic pathways in CMML (Supplementary Table 4). We then asked if those mutations were causally related with the disease by using transcriptional reporter and *in vitro* differentiation assays. We demonstrated that selected mutations had the ability to affect Notch activity negatively either as dominant negative (MAML1Q345X) or null (NCSTNA433T) alleles (Fig. 4c, d and Supplementary Fig. 16). This is the first description, to our knowledge, of somatic Notch pathway loss-of-function mutations in human cancer.

The presented studies identify novel inactivating Notch pathway mutations and suggest that γ SE complex/Notch signalling controls early HSC/MPP commitment decisions in bone marrow. A significant portion of this regulation is controlled by the Hes family of transcriptional repressors. Most importantly, our studies suggest that silencing Notch activity leads to the development of myeloid leukaemia, implying a novel tumour-suppressor function for the Notch pathway in haematopoiesis. Although mutation-mediated pathway silencing can be found in CMML, it is conceivable that there are additional control mechanisms, including epigenetic silencing of Notch pathway target genes. Whatever the mode of regulation, our observations suggest that reversible activation of the Notch pathway may represent an attractive future therapy, targeting specifically the progression and relapse of granulocytic and monocytic neoplasms.

METHODS SUMMARY

Animals. All mice were kept in specific pathogen-free animal facilities at the New York University School of Medicine. Mx1-Cre⁺ animals were injected with 20 μ g polyI:polyC per gram of body weight for a total of six injections. The injections were initiated 14 days after birth and done every 2 days. Animals were analysed 4–6 weeks after the last injection unless indicated otherwise. All animal experiments were done in accordance to the guidelines of the New York University School of Medicine Institutional Animal Care and Use Committee.

Antibodies and fluorescence-activated cell sorting analysis. Antibody staining and fluorescence-activated cell sorting (FACS) were performed as previously described²². All antibodies were purchased from BD-Pharmingen or e-Bioscience. We used the following antibodies: c-kit (2B8), Sca-1 (D7), Mac-1 (M1/70), Gr-1 (RB6-8C5), NK1.1 (PK136), TER-119, CD3 (145-2C11), CD19 (1D3), IL7R α (A7R34), CD34 (RAM34), Fc γ RII/III (2.4G2), Flk-2/Flt-3 (A2F10.1), CD4 (RM4-5), CD4 (H129.19), CD8 (53-6.7), CD45.1 (A20), CD45.2 (104), CD150 (9D1), CD48 (HM481). Bone marrow lineage antibody cocktail included Mac-1, Gr-1, NK1.1, TER-119, CD3 and CD19.

OP9-DL1/DL4 *in vitro* co-culture. OP9-DL1/DL4 cells were maintained in MEM with 20% fetal bovine serum. Ten thousand purified infected (green fluorescent protein (GFP⁺)) LSK or cKit⁺ bone marrow progenitors were seeded into a six-well plate with confluent OP9 cells in the presence of 10 ng ml⁻¹ SCF, 10 ng ml⁻¹ interleukin-3 (IL-3), 5 ng ml⁻¹ Flt3-L, 10 ng ml⁻¹ IL6 and 5 ng ml⁻¹ IL-7. Flow cytometric analysis was performed on an LSRII (BDIS). Haematopoietic cells were gated using CD45 cell-surface expression.

Microarray analysis. LSK or GMP cells from individual mice were used. To generate sufficient sample quantities for oligonucleotide gene chip hybridization, we used the Ovation RNA Amplification System V2 (Nugen) for antisense RNA (cRNA) amplification and labelling. The amplified cRNA was labelled and hybridized to the Mouse 430.2 microarrays (Affymetrix). The data were normalized using the previously published robust multi-array average algorithm using the GeneSpring GX software (Agilent).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 August 2010; accepted 14 March 2011.

- Hurlbut, G. D., Kankel, M. W., Lake, R. J. & Artavanis-Tsakonas, S. Crossing paths with Notch in the hyper-network. *Curr. Opin. Cell Biol.* **19**, 166–175 (2007).
- De Strooper, B. Nicastrin: gatekeeper of the γ -secretase complex. *Cell* **122**, 318–320 (2005).
- Aifantis, I., Raetz, E. & Buonamici, S. Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nature Rev. Immunol.* **8**, 380–390 (2008).
- Robert-Moreno, A. *et al.* Impaired embryonic haematopoiesis yet normal arterial development in the absence of the Notch ligand Jagged1. *EMBO J* **27**, 1886–1895 (2008).
- Zuniga-Pflucker, J. C. T-cell development made simple. *Nature Rev. Immunol.* **4**, 67–72 (2004).

- Rothenberg, E. V. & Taghon, T. Molecular genetics of T cell development. *Annu. Rev. Immunol.* **23**, 601–649 (2005).
- Grabher, C., von Boehmer, H. & Look, A. T. Notch 1 activation in the molecular pathogenesis of T-cell acute lymphoblastic leukaemia. *Nature Rev. Cancer* **6**, 1–13 (2006).
- Kuhn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice. *Science* **269**, 1427–1429 (1995).
- Stadtfeld, M. & Graf, T. Assessing the role of hematopoietic plasticity for endothelial and hepatocyte development by non-invasive lineage tracing. *Development* **132**, 203–213 (2005).
- Emanuel, P. D. Juvenile myelomonocytic leukemia and chronic myelomonocytic leukemia. *Leukemia* **22**, 1335–1342 (2008).
- Challen, G. A., Boles, N. C., Chambers, S. M. & Goodell, M. A. Distinct hematopoietic stem cell subtypes are differentially regulated by TGF- β 1. *Cell Stem Cell* **6**, 265–278 (2010).
- Ng, S. Y., Yoshida, T., Zhang, J. & Georgopoulos, K. Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. *Immunity* **30**, 493–507 (2009).
- Krivtsov, A. V. *et al.* Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **442**, 818–822 (2006).
- Dumortier, A. *et al.* Atopic dermatitis-like disease and associated lethal myeloproliferative disorder arise from loss of notch signaling in the murine skin. *PLoS ONE* **5**, e9258 (2010).
- Radtke, F. *et al.* Deficient T cell fate specification in mice with an induced inactivation of Notch1. *Immunity* **10**, 547–558 (1999).
- Demehri, S. *et al.* Notch-deficient skin induces a lethal systemic B-lymphoproliferative disorder by secreting TSLP, a sentinel for epidermal integrity. *PLoS Biol* **6**, e123 (2008).
- Buonamici, S. *et al.* CCR7 signalling as an essential regulator of CNS infiltration in T-cell leukaemia. *Nature* **459**, 1000–1004 (2009).
- Santaguida, M. *et al.* JunB protects against myeloid malignancies by limiting hematopoietic stem cell proliferation and differentiation without affecting self-renewal. *Cancer Cell* **15**, 341–352 (2009).
- Mercher, T. *et al.* Notch signaling specifies megakaryocyte development from hematopoietic stem cells. *Cell Stem Cell* **3**, 314–326 (2008).
- Taghon, T. N., David, E. S., Zuniga-Pflucker, J. C. & Rothenberg, E. V. Delayed, asynchronous, and reversible T-lineage specification induced by Notch/Delta signaling. *Genes Dev.* **19**, 965–978 (2005).
- Tefferi, A. Novel mutations and their functional and clinical relevance in myeloproliferative neoplasms: JAK2, MPL, TET2, ASXL1, CBL, IDH and IKZF1. *Leukemia* **24**, 1128–1138 (2010).
- Aifantis, I., Feinberg, J., Fehling, H. J., Di Santo, J. P. & von Boehmer, H. Early T cell receptor β gene expression is regulated by the pre-T cell receptor-CD3 complex. *J. Exp. Med.* **190**, 141–144 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Fishel, F. Radtke and R. Kopan for donating mouse strains; P. Lopez and the New York University Flow Facility for cell sorting; and A. Heguy and the Geoffrey Beene Translational Core laboratory for assistance with DNA resequencing. The New York University Cancer Institute Genomics Facility helped with micro-array processing. This work was supported by the National Institutes of Health (R01CA133379, R01CA105129, R21CA141399, R01CA149655 to I.A.; R01CA1328234 to R.L.L. and F.M.; U54CA143798 to F.M.), the Leukemia & Lymphoma Society (to I.A.), the American Cancer Society (to I.A.), the Irma T. Hirsch Trust, the Dana Foundation, The Mallinckrodt Foundation, the Alex's Lemonade Stand Foundation (to I.A.), and the Fund for Scientific Research Flanders (Fonds Wetenschappelijk Onderzoek) and its Odysseus Research Program (to T.T.). A.E. was supported by the National Cancer Institute (1P01CA97403, Project 2) and a gift from the Berrie Foundation. A.K. was supported by an EU Marie Curie International Re-integration Grant. I.v.D.W. was supported by the Institute for the Promotion of Innovation by Science and Technology in Flanders (Agentschap voor Innovatie door Wetenschap en Technologie). S.C. was supported by the Hope Street Kids Foundations and P.O. by the New York University Medical Scientist Training Program. T.T. was supported by the Fonds Wetenschappelijk Onderzoek. O.A.W. was supported by the Clinical Scholars Program at Memorial Sloan Kettering Cancer Center and by the American Society of Hematology. R.L.L. is an Early Career Award recipient of the Howard Hughes Medical Institute and is the Geoffrey Beene Junior Chair at Memorial Sloan Kettering Cancer Center. I.A. is a Howard Hughes Medical Institute Early Career Scientist.

Author Contributions I.A., C.L. and A.K. conceived the study and designed all experiments. A.K. and A.E. helped with experimental planning and generated the Ncstn^{fl/fl} mice. C.L. performed most of the mouse experiments and *in vitro* studies aided by P.O., S.B., S.C., T.T. and E.A. R.L.L., O.A.-W. and M.B. performed and analysed human leukaemia sample exon sequencing. H.H. and F.M. helped with disease modelling and computational analysis of disease progression. I.v.D.W. and T.T. performed the human stem cell differentiation assays. C.L. and S.I. analysed mouse disease pathology. J.Z. processed and analysed gene expression data.

Author Information The microarray data are deposited in Gene Expression Omnibus of the National Center for Biotechnical Information under accession numbers GSE27794, GSE27799 and GSE27811. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to I.A. (iannis.aifantis@nyumc.org) or A.K. (aklinakis@bioacademy.gr).

METHODS

Animals. Genotyping of $N1^{flf} N2^{flf} N3^{-/-}$ mice^{23–26} was performed as previously reported. $Ncstn^{flf} Mx1-Cre^{+}$ and $N1^{flf} N2^{flf} N3^{-/-} Mx1-Cre^{+}$ animals were injected with 20 μ g polyI:polyC per gram of body weight for a total of six injections. The injections were initiated 14 days after birth and done every 2 days. Animals were analysed 4–6 weeks after the last injection unless indicated otherwise. All animal experiments were done in accordance to the guidelines of the New York University School of Medicine Institutional Animal Care and Use Committee.

Generation of the $Ncstn^{flf}$ mice. To generate a conditional $Ncstn$ allele we used standard ES cell targeting approaches. Exons 5–7 of the $Ncstn$ locus were targeted for Cre-dependent recombination. Removal of these exons generates an in-frame stop codon that prematurely terminates translation. As the homology region we used a 7 kb BsrGI–EcoRI genomic fragment spanning introns 2–11. A PGK–NEO cassette and a loxP site were cloned into an EcoRV site located in intron 4, whereas the second loxP site was subcloned into a unique PmlI site of intron 7. Embryonic stem cell colonies (129/sj) were screened by Southern blot of genomic DNA digested with BamHI (Supplementary Fig. 1b). Correctly targeted embryonic stem cells were injected in C57BL/6 blastocysts and chimaeras were generated. After verification of germline transmission, the PGK–NEO cassette was removed using a germline FLP mouse²⁷. $Ncstn^{flf} Flp^{+}$ mice were crossed to the Mx-cre strain. Genotyping primers and probes are available upon request.

Antibodies and FACS analysis. Antibody staining and FACS analysis was performed as previously described²². All antibodies were purchased from BD-Pharmingen or e-Bioscience. We used the following antibodies: c-kit (2B8), Sca-1 (D7), Mac-1 (M1/70), Gr-1 (RB6-8C5), NK1.1 (PK136), TER-119, CD3 (145-2C11), CD19 (1D3), IL-7 α (A7R34), CD34 (RAM34), Fc γ RII/III (2.4G2), Flk-2/Flt-3 (A2F10.1), CD4 (RM4-5), CD4 (H129.19), CD8 (53-6.7), CD45.1 (A20), CD45.2 (104), CD150 (9D1), CD48 (HM481). Bone marrow lineage antibody cocktail included Mac-1, Gr-1, NK1.1, TER-119, CD3, CD19. For western blotting, goat polyclonal anti-Nicastrin antibody (N19, sc-14369, Santa Cruz) and mouse monoclonal anti-Actin antibody (Clone C4, MAB1501R, Millipore) were used. Annexin V/7AAD staining was done using an annexin V PE detection kit (BD Pharmingen, 559763) following the manufacturer's protocol.

PCR with reverse transcription. Total RNA was isolated using the RNeasy Plus Mini Kit (Qiagen) and cDNA was synthesized using the SuperScript First-Strand Kit (Invitrogen). Quantitative PCR was performed using SYBR green iMaster and a LightCycler 480 (Roche) with the following primer sequences (melting temperature = 60 °C for all primers): *Cebpa* forward TTACAACAGGCCAGGT TTCC, *Cebpa* reverse CTCTGGGATGGATCGATTGT; *Pu.1* forward ATGGA AGGGTTTTCCTCACCGCC, *Pu.1* reverse GTCCACGCTCTGCAGCTCTGT GAA; *Gata2* forward AACGCTGTGGCCTCTACTA, *Gata2* reverse TCT CTTGCATGCACTTGGAG; *Cebpd* forward ATCGCTGCAGCTTCTATGT, *Cebpd* reverse AGTCATGCTTCCCGTGTTC; *Hes1* forward TCCAAGCTAG AGAAGGCAGAC, *Hes1* reverse TGATCTGGGTCATGCAGTTG; *Gata1* forward ACTGTGGAGCAACGGCTACT, *Gata1* reverse TCCGCCAGAGTGTG GTAGTG; *Ncstn* forward CTGGCGCTGCACTGTATGAG, *Ncstn* reverse GGAGACGGCGATGTAGTGTGAAG.

Bone marrow transplantation assays. Five hundred thousand bone marrow cells ($Ly5.2^{+}$) were transplanted by retro-orbital intravenous injections into lethally irradiated (960 cGy) BL6SJL ($Ly5.1^{+}$) recipient mice. Four weeks after transplantation, mice were injected with 20 μ g polyI:polyC per gram of body weight for a total of six injections. Peripheral blood of recipient mice was collected at 4, 7, 9 and 12 weeks after transplantation. Recipient mice were killed 16 weeks after transplantation for analysis.

Microarray analysis. LSK or GMP cells from individual mice were used. Microarray analysis was performed as previously described²⁸. Briefly, freshly isolated cells were sorted by surface marker expression, and total RNA was extracted using the RNeasy kit (Qiagen). To generate sufficient sample quantities for oligonucleotide gene chip hybridization, we used the Ovation RNA Amplification System V2 (Nugen) for cRNA amplification and labelling. The amplified cRNA was labelled and hybridized to the Mouse 430.2 microarrays (Affymetrix). The Affymetrix gene expression profiling data were normalized using the previously published robust multi-array average algorithm using the GeneSpring GX software (Agilent). The gene-expression intensity presentations were generated with Matrix2png software (<http://chibi.ubc.ca/matrix2png/bin/matrix2png.cgi>) or Multi Experiment Viewer software (<http://www.tm4.org/mev/>).

Gene set enrichment analysis. Gene set enrichment analysis was performed using Gene Set Enrichment Analysis software^{29,30} (<http://www.broadinstitute.org/gsea/>) using gene set as permutation type, 1,000 permutations and \log_2 ratio of classes as metric for ranking genes.

The 'myeloid signature' gene set was generated using a systematic approach based on the comparison of gene expression arrays from WT LSK and WT GMP. Genes that were significantly upregulated in GMP compared with LSK

(over 1.5-fold induction $P < 0.05$) were used to define myeloid signature genes. The list was trimmed of genes of unknown function and those related to metabolism.

Other myeloid-specific gene sets used in the analysis were taken from gene sets already present in the MSig database of the Broad Institute.

Retroviral infection of Lineage^{neg}cKit⁺ bone marrow cells or LSK cells. Bone marrow cells were enriched for cKit-positive cells using either the EasySep kit (StemCell Technology) or the Dynabead kit (DYNAL, Invitrogen). LSK cells were flow sorted using lineage markers Mac-1, Gr-1, NK1.1, TER-119, CD3, CD19, Sca1 and cKit. Cells were subsequently cultured in OPTI-MEM supplemented with 10% fetal bovine serum, 100 μ g ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin, 50 ng ml⁻¹ SCF and Flt3l, and 10 ng ml⁻¹ IL-6 and IL-7. For retroviral production, Plat-E cells were transfected with appropriate retroviral expression constructs by the calcium phosphate method. Virus supernatant was collected 48 h after transfection and used directly for spin infection of cKit positive-enriched bone marrow cells or sorted LSK cells at 660g for 90 min. Forty-eight hours after infection, lineage-negative GFP-positive cells were sorted for RT-PCR analysis, methylcellulose plating assays or OP9 co-culture assay.

In vitro differentiation assays. Total bone marrow (15,000), sorted LSK (500) and GMP (500) cells were plated in triplicates into cytokine-supplemented methylcellulose medium (MethoCult 3434, Stem Cell Technologies). Colony type was scored after 10 days of culture. Replating was performed after 8 days of culture. For experiments involving Hes1 overexpression, 10,000 GFP⁺ infected cells were plated in duplicate on 35 mm cytokine-supplemented methylcellulose medium (MethoCult 3434, Stem Cell Technologies). Cells were recovered 8 days later, stained and analysed by FACS as described.

Human progenitor/OP9 co-cultures. Bone marrow samples were obtained and used according to the guidelines of the Medical Ethical Commission of Ghent University Hospital (Belgium). Bone marrow mononuclear cells were isolated through Lymphoprep density-gradient centrifugation and CD45⁺ glycophorin-A positive (Gly-A⁺) cells were enriched by micro-magnetic beads (Miltenyi Biotec). Subsequently, CD45⁺ Gly-A⁺ cells were labelled with CD34-FITC, CD38-PE, CD19-APC and CD14-APC to sort CD34⁺CD38⁻Lin⁻ cells on a FACSaria to a purity greater than 98%. OP9-control and OP9 cells expressing human ligands Dll1, Dll4, Jag1 and Jag2 were generated by retroviral transduction of OP9 cells (provided by J. C. Zúñiga-Pflücker), followed by sorting for eGFP⁺ cells, as described previously³¹. Purified bone-marrow progenitors (2.0×10^3 – 2.6×10^3) were seeded into a 24-well plate with confluent OP9 cells in the presence of 20 ng ml⁻¹ SCF, 20 ng ml⁻¹ Flt-3L, 20 ng ml⁻¹ TPO, 10 ng ml⁻¹ GM-CSF and 10 ng ml⁻¹ G-CSF. Co-cultures were performed and prepared for analysis as described previously³². Flow cytometric analysis was performed on an LSRII (BDIS). Human cells were gated using a CD45 monoclonal antibody and dead cells were excluded using propidium iodide.

OP9-DL1/DL4 in vitro co-culture. OP9 cells were maintained in MEM with 20% fetal bovine serum. Ten thousand purified infected (GFP⁺) LSK or cKit⁺ bone marrow progenitors were seeded into a six-well plate with confluent OP9 cells in the presence of 10 ng ml⁻¹ SCF, 10 ng ml⁻¹ IL-3, 10 ng ml⁻¹ IL-6, 5 ng ml⁻¹ Flt3-L and 5 ng ml⁻¹ IL-7. Flow cytometric analysis was performed on an LSRII (BDIS). Haematopoietic cells were gated using CD45 cell-surface expression and GFP expression.

Chromatin immunoprecipitation assays. Chromatin immunoprecipitation assays were performed with 80 μ g of genomic DNA from 32D-MIG- or 32D-HES1-infected cells following standard procedures. Briefly, chromatin was cross-linked with 1% formaldehyde, and sheared by sonication. Immunoprecipitation was performed with immunoglobulin-G or anti-HES1, followed by incubation with protein A magnetic beads/salmon sperm DNA (Invitrogen Dynal). DNA isolated from antibody bound fraction was eluted after washing, extracted with phenol/chloroform and precipitated with ethanol. Real-time PCR quantification of immunoprecipitated DNA was performed with the SYBR Green PCR Master Mix (Roche) and primers designed to amplify regions covering each putative HES1 binding site in the *Pu.1* and *Cebpa* promoters: *Pu.1-1* forward ACGTTCAAGGGTTGGAGAAA, *Pu.1-1* reverse GCCAATTAGGGCCAACAGTA; *Pu.1-2* forward GACCAAAGTCTT CCACCTGA, *Pu.1-2* reverse CTGGGAGGGAGAAAGGCTA; *Cebpa-1* forward CCAAAGCAGTCTCCCAACCTC, *Cebpa-1* reverse CCACTCCAGCCAACAC TA; *Cebpa-2* forward CGCCTAACCACGACCAC, *Cebpa-2* reverse AGTAG GATGGTGCCTGCTG.

Histological analyses. Mice were killed and autopsied, then dissected tissue samples or tumours were fixed for 24 h in 10% buffered formalin, dehydrated and embedded in paraffin. Paraffin blocks were sectioned at 5 μ m and stained with haematoxylin and eosin.

Luciferase reporter assays. pGL3 promoter reporter plasmid (Promega) was used to clone *Pu.1*, *Cebpa* or *Hes1* promoter sequences. pGL3 promoter empty vector or the vector containing *Pu.1*, *Cebpa* or *Hes1* promoter sequences, HA-HES1 expression plasmid or pCDNA3.1 control plasmid and *Renilla* expression vector

were co-transfected into HEK293T cells using regular calcium phosphate transfection protocol. Luciferase activities were examined 24 h after transfection using the Dual Luciferase reporter assay system (Promega) following the manufacturer's instructions, and normalized to *Renilla* activity.

Statistical analysis. The means of each data set were analysed using Student's *t*-test, with a two-tailed distribution and assuming equal sample variance.

23. Yang, X. *et al.* Notch activation induces apoptosis in neural progenitor cells through a p53-dependent pathway. *Dev. Biol.* **269**, 81–94 (2004).
24. Saito, T. *et al.* Notch2 is preferentially expressed in mature B cells and indispensable for marginal zone B lineage development. *Immunity* **18**, 675–685 (2003).
25. Mitchell, K. J. *et al.* Functional analysis of secreted and transmembrane proteins critical to mouse development. *Nature Genet.* **28**, 241–249 (2001).
26. Leighton, P. A. *et al.* Defining brain wiring patterns and mechanisms through gene trapping in mice. *Nature* **410**, 174–179 (2001).
27. Rodriguez, C. I. *et al.* High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. *Nature Genet.* **25**, 139–140 10.1038/75973 (2000).
28. Thompson, B. J. *et al.* Control of hematopoietic stem cell quiescence by the E3 ubiquitin ligase Fbw7. *J. Exp. Med.* **205**, 1395–1408 (2008).
29. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
30. Mootha, V. K. *et al.* PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34**, 267–273 (2003).
31. Van de Walle, I. *et al.* An early decrease in Notch activation is required for human TCR- $\alpha\beta$ lineage differentiation at the expense of TCR- $\gamma\delta$ T cells. *Blood* **113**, 2988–2998 (2009).
32. Taghon, T. *et al.* Notch signaling is required for proliferation but not for differentiation at a well-defined β -selection checkpoint during human T-cell development. *Blood* **113**, 3254–3263 (2009).

Acetylation-dependent regulation of endothelial Notch signalling by the SIRT1 deacetylase

Virginia Guarani¹, Gianluca Deflorian^{2*}, Claudio A. Franco^{3*}, Marcus Krüger⁴, Li-Kun Phng^{3†}, Katie Bentley³, Louise Toussaint⁵, Franck Dequiedt⁶, Raul Mostoslavsky⁶, Mirko H. H. Schmidt⁷, Barbara Zimmermann¹, Ralf P. Brandes⁸, Marina Mione², Christoph H. Westphal⁹, Thomas Braun⁴, Andreas M. Zeiher¹⁰, Holger Gerhardt^{3,11}, Stefanie Dimmeler¹ & Michael Potente^{1,10}

Notch signalling is a key intercellular communication mechanism that is essential for cell specification and tissue patterning, and which coordinates critical steps of blood vessel growth^{1–3}. Although subtle alterations in Notch activity suffice to elicit profound differences in endothelial behaviour and blood vessel formation^{2,3}, little is known about the regulation and adaptation of endothelial Notch responses. Here we report that the NAD⁺-dependent deacetylase SIRT1 acts as an intrinsic negative modulator of Notch signalling in endothelial cells. We show that acetylation of the Notch1 intracellular domain (NICD) on conserved lysines controls the amplitude and duration of Notch responses by altering NICD protein turnover. SIRT1 associates with NICD and functions as a NICD deacetylase, which opposes the acetylation-induced NICD stabilization. Consequently, endothelial cells lacking SIRT1 activity are sensitized to Notch signalling, resulting in impaired growth, sprout elongation and enhanced Notch target gene expression in response to DLL4 stimulation, thereby promoting a non-sprouting, stalk-cell-like phenotype. *In vivo*, inactivation of Sirt1 in zebrafish and mice causes reduced vascular branching and density as a consequence of enhanced Notch signalling. Our findings identify reversible acetylation of the NICD as a molecular mechanism to adapt the dynamics of Notch signalling, and indicate that SIRT1 acts as rheostat to fine-tune endothelial Notch responses.

We investigated the role of SIRT1 in the regulation of endothelial Notch signalling. Notch activity was analysed in response to DLL4 stimulation in scrambled or *SIRT1* short interfering (si)RNA-transfected human umbilical vein endothelial cells by assessing the activity of the Notch reporter genes TP1-luciferase and 4×CBF1-luciferase as well as the expression of the endogenous Notch target genes *NRARP* and *HEY2*. Knockdown of SIRT1 by siRNA did not alter basal Notch activity, but significantly enhanced Notch reporter gene activity and target gene expression after DLL4 stimulation (Fig. 1a–c and Supplementary Fig. 2a, b). Comparable results were obtained with a non-related pool of *SIRT1* siRNAs (Supplementary Fig. 2b, c). Importantly, enhanced Notch responsiveness to DLL4 in *SIRT1*-deficient endothelial cells was abrogated by the γ -secretase inhibitor dibenzazepine (DBZ) or by mutation of the CBF1-binding sites in the Notch reporter gene (Fig. 1a, b). Conversely, activation of SIRT1 signalling using the small molecule SRT2183 (ref. 4) inhibited DLL4-mediated induction of Notch reporter activity and target gene expression, indicating that SIRT1 negatively modulates DLL4/Notch signalling in endothelial cells (Fig. 1d, e).

Analysis of the basal messenger RNA expression of critical Notch pathway components such as *NOTCH1*, *NOTCH4*, *MAML1* and

NCOR1 in control and *SIRT1*-deficient endothelial cells revealed no significant differences (Supplementary Fig. 2d). However, *SIRT1*-deficient endothelial cells displayed markedly enhanced Notch activity after overexpression of NICD, as assessed using different Notch reporter genes (Fig. 1f and Supplementary Fig. 2e, f). On the other hand, co-transfection of SIRT1 with NICD inhibited Notch activity in a dose-dependent manner (Supplementary Fig. 2g). The inhibitory effect of SIRT1 required its deacetylase activity, as a catalytically inactive SIRT1 mutant (SIRT1 H363Y) was unable to inhibit Notch responses (Fig. 1g). Taken together, these results demonstrate that inhibition of Notch signalling by SIRT1 occurs in endothelial cells that receive activating DLL4/Notch signals and indicate a regulatory mechanism involving NICD as well as the deacetylase activity of SIRT1.

Co-immunoprecipitation experiments illustrated that overexpressed NICD associates with SIRT1–Flag and endogenous SIRT1, but not with other tested sirtuins (Fig. 1h and Supplementary Figs 3a–c). Intriguingly, NICD–V5 levels were consistently lower in SIRT1–Flag overexpressing lysates (Fig. 1h and Supplementary Fig. 3a). This reduction in NICD protein levels was not observed when SIRT1 was co-expressed with a NICD mutant lacking the carboxy-terminal PEST domain important for proteasomal degradation (NICD(Δ C)), although its binding to SIRT1 was retained (Fig. 1i). These data indicate that SIRT1 might negatively regulate Notch signalling by promoting NICD degradation. Consistent with this model, the interaction between endogenous NICD and SIRT1 was enhanced when endothelial cells were treated with the proteasomal inhibitor MG132 (Fig. 1j). Further characterization of this protein interaction revealed that deletion of the SIRT1 catalytic domain abolished the binding of SIRT1 to NICD, indicating that NICD might be a substrate for the deacetylase SIRT1 (Supplementary Fig. 3d, e).

To assess whether SIRT1 acts as a NICD deacetylase, we asked whether NICD is an acetylated protein. We transfected 293 cells with NICD–V5 and the acetyltransferases PCAF and p300, which bind to NICD and function as coactivators at Notch-regulated promoters^{5,6}. NICD was robustly acetylated in cells co-expressing PCAF or p300 (Supplementary Fig. 4a, b). In contrast, Tip60, which has been reported to acetylate NICD after ultraviolet radiation⁷, was unable to acetylate NICD under our experimental conditions (Supplementary Fig. 4c). Blocking deacetylase activity with the sirtuin inhibitor nicotinamide (NAM) induced acetylation of overexpressed as well as endogenous NICD in endothelial cells, whereas the class I/II histone deacetylase inhibitor trichostatin A (TSA) had only a minor effect (Fig. 1k, l). Notably, knockdown of SIRT1 increased basal and p300-induced NICD acetylation, whereas overexpression prevented its acetylation

¹Institute for Cardiovascular Regeneration, Centre of Molecular Medicine, Goethe University, D-60590 Frankfurt, Germany. ²IFOM, the FIRC Institute of Molecular Oncology, IFOM-IEO Campus, 20139 Milan, Italy. ³Vascular Biology Laboratory, London Research Institute – Cancer Research UK, WC2A 3LY London, UK. ⁴Max Planck Institute for Heart and Lung Research, Department of Cardiac Development and Remodeling, D-61231 Bad Nauheim, Germany. ⁵Laboratory of Protein Signaling and Interactions, GxABT, B-5030 Gembloux and Interdisciplinary Cluster for Applied Genoproteomics (GIGA-R), University of Liege, B-4000 Sart-Tilman, Belgium. ⁶Massachusetts General Hospital Cancer Center, Harvard Medical School, Boston, Massachusetts 02114, USA. ⁷Molecular Signal Transduction, Institute of Neurology (Edinger Institute), Goethe University, D-60590 Frankfurt, Germany. ⁸Vascular Research Centre, Institute for Cardiovascular Physiology, Goethe University, D-60590 Frankfurt, Germany. ⁹Sirtis, a GSK Company, Cambridge, Massachusetts 02139, USA. ¹⁰Department of Cardiology, Internal Medicine III, Goethe University, D-60590 Frankfurt, Germany. ¹¹Consultant Group Leader, Vascular Patterning Laboratory, Vesalius Research Center, VIB, Campus Gasthuisberg, B-3000 Leuven, Belgium. [†]Present address: Cell Biology and Biophysics, European Molecular Biology Laboratory, D-69117 Heidelberg, Germany.

*These authors contributed equally to this work.

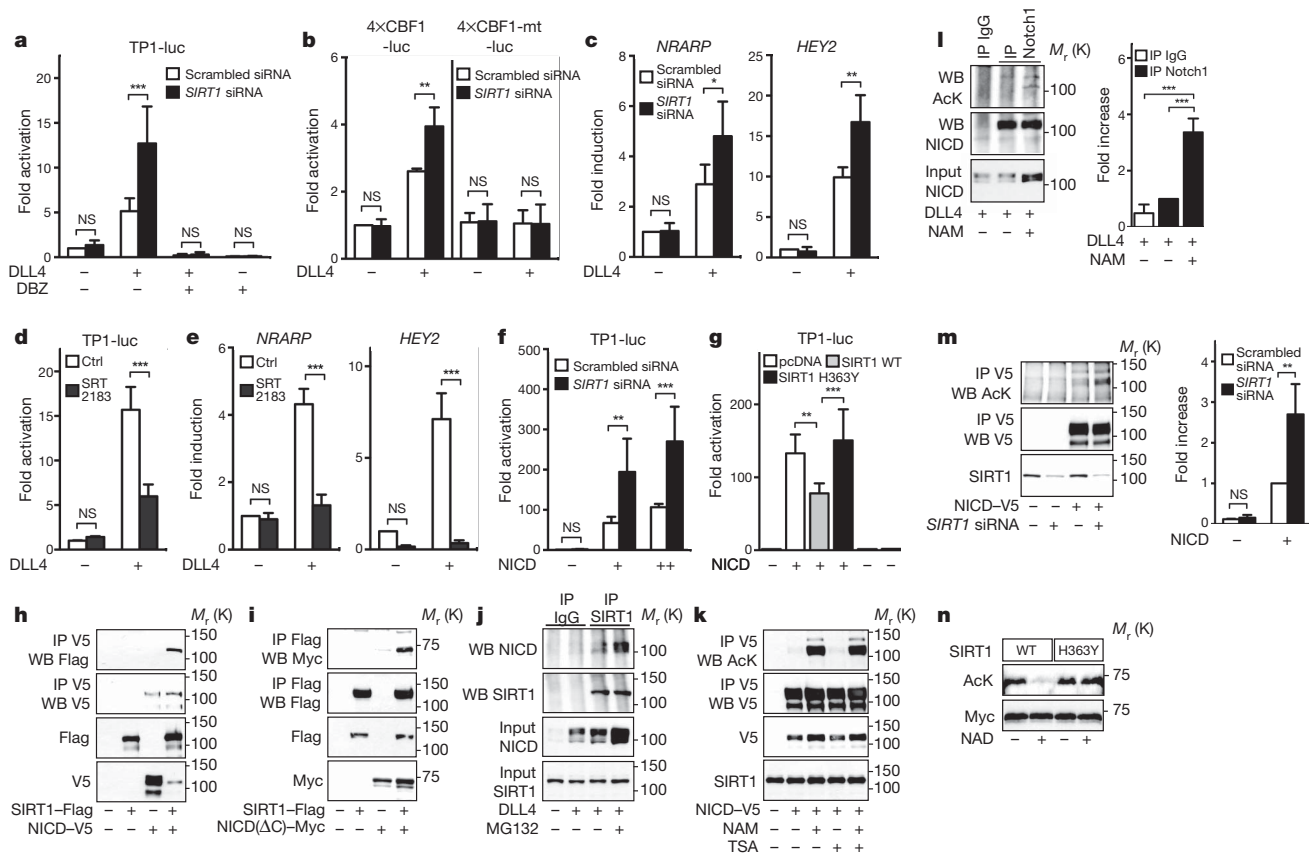


Figure 1 | SIRT1 limits endothelial DLL4/Notch signalling and targets NICD for deacetylation. **a, b**, TP1- (**a**), 4×CBF1- or 4×CBF1 mutant- (mt) luciferase activity (**b**) in control (scrambled) or *SIRT1*-siRNA-transfected endothelial cells. **c**, *NRARP* and *HEY2* mRNA levels in scrambled or *SIRT1* siRNA-transfected endothelial cells. **d, e**, TP1-luciferase activity (**d**) and *NRARP* and *HEY2* mRNA levels (**e**) in endothelial cells stimulated with SRT2183. **f**, TP1-luciferase activity in control and *SIRT1*-deficient endothelial cells co-transfected with NICD. ++, increasing amounts of NICD. **g**, TP1-luciferase activity in endothelial cells transfected with combinations of NICD, *SIRT1* and *SIRT1* H363Y. **h, i**, Co-immunoprecipitations from *SIRT1*-Flag and NICD-V5 or

(Fig. 1m and Supplementary Fig. 4d, e). Moreover, wild type but not the inactive H363Y SIRT1 mutant deacetylated NICD (NICD(Δ C)) in an *in vitro* deacetylation assay in a NAD^+ -dependent manner (Fig. 1n). These data demonstrate that NICD is reversibly acetylated and identify SIRT1 as a bona fide NICD deacetylase.

Using liquid chromatography-tandem mass spectrometry analysis (LC-MS/MS) we found 14 acetylation sites in the NICD targeting conserved lysine residues (Fig. 2a, b and Supplementary Fig. 5a, b). Mutation of these lysines to arginine (NICD(14KR)) abolished NICD acetylation induced by PCAF, p300, or upon NAM treatment or SIRT1 knockdown (Fig. 2c and Supplementary Fig. 5c-e). Compared to wild-type NICD, the NICD(14KR) mutant had a similar ability to activate Notch reporter genes, but was resistant to changes in the level of SIRT1 (Fig. 2d and Supplementary Fig. 6a, b). These data indicate that a set of NICD lysine residues confer negative regulation of Notch signalling by SIRT1.

Intriguingly, conditions that favour NICD acetylation increased NICD protein levels (Fig. 1k, l and Supplementary Figs 4a, b and 7a). Because NICD undergoes rapid ubiquitin-mediated degradation¹, and acetylation can impair ubiquitination, we explored the effects of SIRT1 on NICD protein degradation and stability. Knockdown of SIRT1 increased, whereas overexpression decreased, NICD protein levels (Fig. 2e, f and Supplementary Fig. 7b). Importantly, protein levels of the NICD(14KR) mutant were not enhanced in SIRT1-deficient or NAM-treated cells (Fig. 2f and Supplementary Fig. 7c).

NICD(Δ)–Myc co-transfected 293 cells. IP, immunoprecipitation; WB, western blot. **j**, Co-immunoprecipitation of NICD and SIRT1 in endothelial cells treated with or without MG132. **k**, Acetylation of NICD–V5 in 293 cells treated with NAM, TSA, or combinations thereof. AcK, acetyl-lysine. **l**, Acetylation of NICD in endothelial cells treated with NAM. **m**, Acetylation of NICD–V5 in *SIRT1*-siRNA-transfected 293 cells. **n**, Deacetylation assay with NICD(Δ)–Myc, recombinant wild-type or H363Y SIRT1. Relative quantifications of NICD acetylation are shown on the right of the panel in **l**, **m**. DLL4 was used to stimulate Notch signalling in endothelial cells in **a–e**, **j** and **l**. All experiments $n \geq 3$; error bars, mean \pm s.d. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; NS, not significant.

Similarly, blocking SIRT1 activity in endothelial cells by NAM or siRNA increased endogenous NICD protein levels, whereas activation of SIRT1 by SRT2183 reduced endogenous NICD protein levels (Fig. 2g, h and Supplementary Fig. 7d). Treatment of endothelial cells with MG132 led to an increase in NICD protein levels, which were not further enhanced by NAM treatment (Supplementary Fig. 7e). MG132 also prevented downregulation of NICD after stimulation with SRT2183 (Fig. 2i). Together with the unaltered *NOTCH1* mRNA expression these data indicate that SIRT1 affects the proteasomal degradation of NICD (Supplementary Fig. 7f). Next, we examined the effect of SIRT1 on the decay of NICD protein levels in response to DLL4 stimulation in cycloheximide-treated endothelial cells. Blocking SIRT1 activity by NAM or *SIRT1* siRNA increased, whereas activation by SRT2183 decreased, the amplitude and duration of endogenous NICD protein levels (Fig. 2j and Supplementary Fig. 8a–d). Concomitantly, SIRT1 inhibition by NAM or siRNA-mediated knock-down reduced levels of ubiquitinated NICD in cells that co-expressed NICD-V5 and HA-tagged ubiquitin (Supplementary Fig. 9a–c). Together, these data support a model in which acetylation interferes with ubiquitin-dependent proteasomal degradation of NICD, hence enhancing Notch responses. The deacetylase activity of SIRT1 counteracts the stabilization of NICD by priming it for ubiquitin-mediated proteolysis (Supplementary Fig. 1).

To assess whether SIRT1 regulates Notch responses *in vivo*, we inactivated the catalytic domain of SIRT1 in endothelial cells

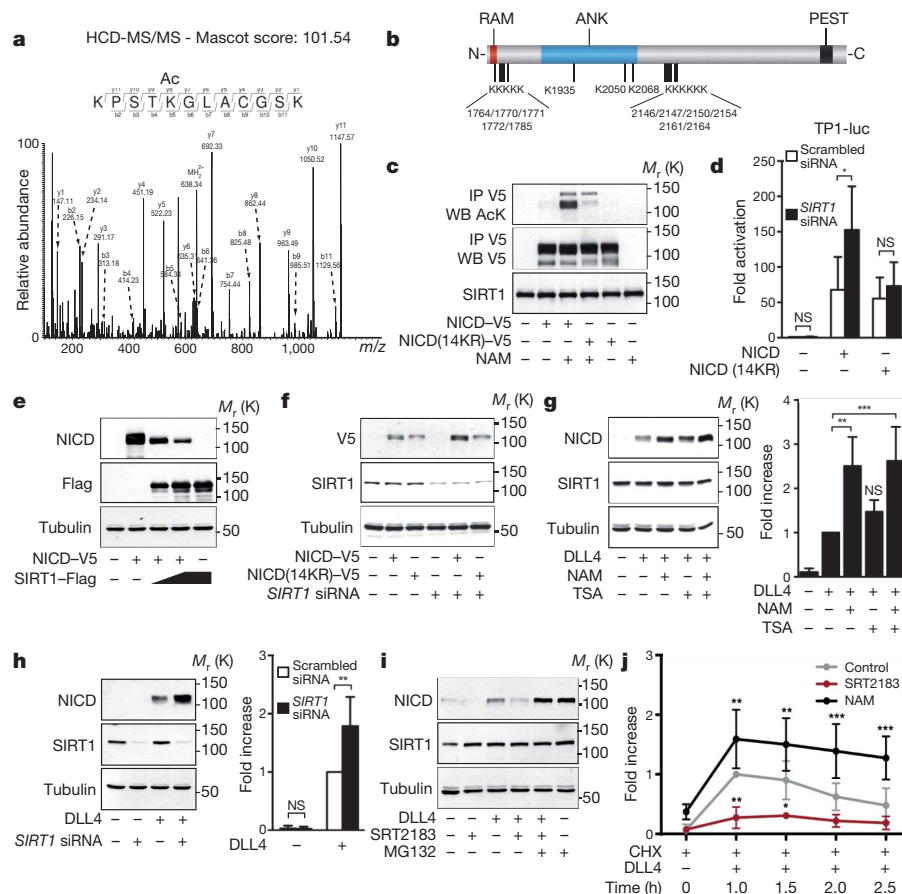


Figure 2 | Destabilization of NICD by SIRT1. **a**, HCD (higher energy C-trap dissociation) MS/MS spectrum showing acetylation of NICD lysine residue 2154. **b**, Overview of acetylated lysines in the NICD. ANK, ankyrin repeats; PEST, proline (P), glutamic acid (E), serine (S) and threonine (T) enriched sequence; RAM, RBPj-associated molecule. **c**, Acetylation of NICD-V5 wild type or 14KR in transfected 293 cells treated with NAM. **d**, TP1-luciferase activity in scrambled and *SIRT1*-siRNA-transfected endothelial cells co-transfected with wild-type or NICD(14KR). **e**, NICD protein levels in 293 cells expressing NICD-V5 and SIRT1-Flag. **f**, NICD-V5 protein levels in 293 cells transfected with scrambled or *SIRT1* siRNA and NICD-V5 wild type or 14KR.

(Supplementary Fig. 10a)⁸ and assessed postnatal retinal angiogenesis in mice. Recent studies showed that DLL4/Notch signalling coordinates retinal angiogenesis by controlling the specification of endothelial cells into tip and stalk cells^{9–14}. Tip cells express high levels of DLL4 and guide nascent sprouts, whereas following stalk cells receive DLL4/Notch signals from tip cells to form the vascular tube. Immunofluorescence staining revealed that Sirt1 is robustly expressed at the angiogenic front of the vasculature, particularly in stalk cells (Fig. 3a, b), in which Notch activity is most prominent^{9,13}. Compared to *Sirt1*^{fllox/+} and *Tie2cre;Sirt1*^{fllox/+} control mice, endothelial-cell-restricted *Sirt1* mutant mice (*Tie2cre;Sirt1*^{fllox/-}) displayed a delayed expansion of the vascular plexus and a significant reduction in vessel density and endothelial cell proliferation (Fig. 3c–j and Supplementary Fig. 10b–e). These phenotypic changes resemble those caused by enhanced Notch activity in stalk cells¹³ and were mirrored by enhanced expression of the Notch target genes *Nrarp*, *Hey2* and *Lfng* in DLL4-treated endothelial cells derived from these mice (Fig. 3k). However, the ability of tip cells to extend filopodia was unaffected in *Sirt1* mutant mice (Supplementary Fig. 10f–i). Together with the prominent expression of Sirt1 in stalk cells, these findings indicate that Sirt1 negatively modulates Notch signalling in stalk cells to facilitate endothelial branching and proliferation, thereby controlling vascular growth and density.

To understand further the effects of SIRT1 on the dynamics of tip/stalk cell behaviour, we used computational modelling¹⁵. Modelling

g, h, NICD protein expression in endothelial cells treated with NAM and/or TSA (**g**) or transfected with scrambled or *SIRT1* siRNA (**h**). Relative quantifications of NICD protein levels are shown on the right of the panel. **i**, NICD protein levels in endothelial cells pre-treated with or without MG132 and/or SRT2183. **j**, Endothelial cells were pre-treated with CHX and NAM or SRT2183 and NICD protein levels analysed at the indicated time points. DLL4 was used to stimulate Notch signalling in endothelial cells in **g–j**. NICD protein levels in **g–i** were assessed after 6 h of DLL4 stimulation. All experiments $n \geq 3$; error bars, mean \pm s.d. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; NS, not significant.

the increase in NICD levels induced by *Sirt1*-deficiency revealed a delayed negative feedback regulation of the DLL4/Notch pathway, resulting in a slower selection rate of tip and stalk cells. The tip cells were, however, phenotypically normal, producing similar numbers of filopodia (Supplementary Movies 1 and 2). The predicted delay in tip and stalk cell selection would translate into both reduced branching and delayed expansion of the vascular plexus, as observed *in vivo*.

Next, we tested whether Notch inhibition would restore endothelial cell responses in *Sirt1* mutants. Treatment of control animals with the γ -secretase inhibitor *N*-[*N*-(3,5-difluorophenacetyl)-*L*-alanyl]-*S*-phenylglycine *t*-butyl ester (DAPT) caused a hyperdense and hyperbranched vessel network, as reported^{9,11,13,14} (Fig. 3l and Supplementary Fig. 10m). In response to DAPT treatment, vessel density in *Sirt1* mutants increased to a similar extent as in DAPT-treated controls, indicating that the reduction in vessel density is a consequence of increased Notch activity (Fig. 3l and Supplementary Fig. 10j–m). Likewise, knockdown of *sirt1* in zebrafish by antisense morpholino-modified oligonucleotides resulted in defects of the trunk vasculature similar to those caused by increased Notch activity in stalk cells^{13,16,17} and were characterized by thin, misguided, poorly connected and hypocellular intersomitic vessels⁸ (Fig. 4a, b, e and Supplementary Fig. 11a, b). Treatment of *sirt1* morphants with DAPT or knockdown of the artery-specific Notch ligand *dll4*, at least in part, normalized the patterning and cellularity of intersomitic vessels

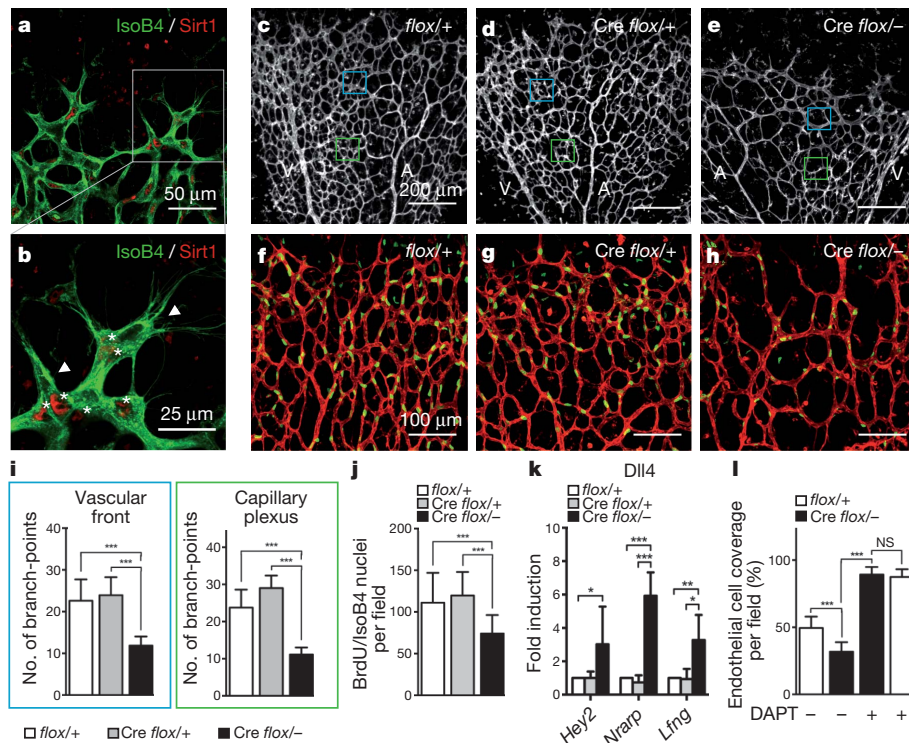


Figure 3 | Inactivation of SIRT1 enhances endothelial Notch responses in mice. **a**, Sirt1 localization (red) in the retinal endothelium (isolectin B4, green). **b**, Higher magnification of inset in **a** is shown. Asterisks indicate Sirt1 expression in stalk endothelial cells; arrowheads indicate some tip cells with weak or absent Sirt1 expression. **c–e**, Images of P5 *Sirt1* control and mutant retinas stained with IsoB4. Blue and green boxes indicate the vascular front and capillary plexus, respectively. A, arteries; V, veins. **f–h**, IsoB4 (red) and BrdU (green) labelling in P5 retinas of the respective genotypes. **i, j**, Statistical summary of the number of vessel branch points and the number of BrdU/IsoB4-positive cells of the respective genotypes. **k**, qPCR of *Hey2*, *Nrarp* and *Lfng* mRNA expression in DLL4-stimulated endothelial cells derived from mice of the respective genotypes. **l**, Statistical summary of the percentage of IsoB4-positive vessel coverage in vehicle and DAPT-treated retinas of control and *Sirt1* mutant mice. All experiments $n \geq 4$; error bars, mean \pm s.d. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; NS, not significant.

(Fig. 4a–e and Supplementary Fig. 11a–c), indicating that misregulation of Notch activity in endothelial cells contributes to the vascular patterning defects in *sirt1* morphants.

To dissect further the effects of SIRT1 on Notch-controlled cell responses, we monitored sprout formation and proliferation in isolated human endothelial cells. Control endothelial cell spheroids extended multicellular sprouts (Supplementary Fig. 12a, b). In contrast, knockdown of SIRT1 impaired sprout formation and elongation⁸ (Supplementary Fig. 12a, b). Importantly, these defects were restored by γ -secretase inhibition (DBZ) (Supplementary Fig. 12a, b). Likewise,

inhibition of endothelial cell growth by DLL4/Notch signalling was enhanced in SIRT1-deficient endothelial cells and rescued by DBZ (Supplementary Fig. 12c). To test the cell-autonomous effect of Sirt1 in modulating Notch responses, we used clusters of differentiating mouse embryonic stem cells (embryoid bodies (EBs)), which form branched vascular networks in response to VEGF-A stimulation¹⁸. Control EBs expressing the fluorophore DsRed-MST (DsRed) developed extensively branched vessel networks, whereas genetic inactivation of *Sirt1* (*Sirt1* ^{$\Delta ex4/\Delta ex4$}) impaired vascular outgrowth and density (Fig. 4f, g). Recent work on genetically mosaic EBs illustrated that

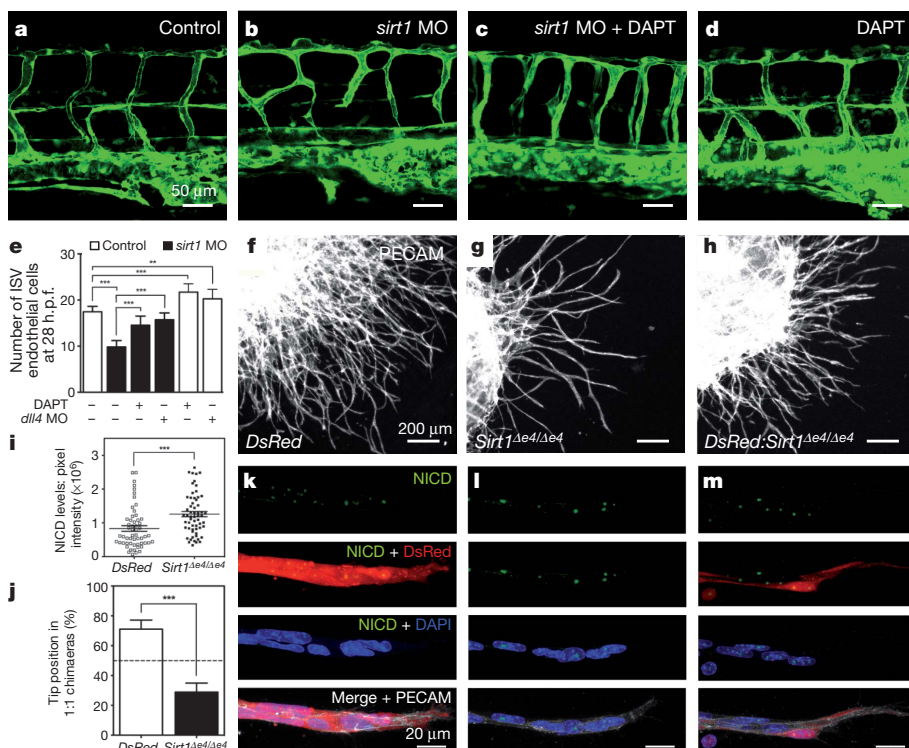


Figure 4 | Inactivation of SIRT1 leads to a cell-autonomous increase in Notch signalling and defective endothelial cell sprouting. **a–d**, Lateral trunk views of 48 h post fertilization *Tg(fli1a:EGFP)y1* zebrafish embryos. Control embryos or *sirt1* morphants were treated with DAPT and DMSO. **e**, Endothelial cell nuclei counts in the intersomitic vessels (ISVs) of transgenic *Tg(fli1a:nEGFP)y7* embryos at 28 h post fertilization (h.p.f.) treated with or without DAPT, or injected with *dll4*-specific morpholinos. **f–h**, Overview of vascular sprouts from EBs of *DsRed* (WT) (**f**), *Sirt1* ^{$\Delta ex4/\Delta ex4$} (**g**) ES cells and 1:1 chimeras (**h**). **i**, Quantification of total pixel intensity of the NICD immunostaining per nucleus in *DsRed* and *Sirt1* mutant endothelial cells. **j**, Quantification of tip-cell contribution of each ES cell genotype in vascular sprouts. **k–m**, NICD immunostaining (green) of individual sprouts of EBs from the respective genotypes. *DsRed* cells (red), *Sirt1* mutant cells (non-labelled), DAPI (blue), PECAM1 (grey). All experiments $n \geq 4$. Error bars, mean \pm s.d. **, $P < 0.01$; ***, $P < 0.001$.

the ability of individual cells to gain the tip position is dynamically and competitively regulated by Dll4/Notch signalling¹⁹. Adding DsRed cells to *Sirt1* mutant cells in a 1:1 ratio partially restored sprouting and branching (Fig. 4f–h), as DsRed cells (wild type for Sirt1) effectively formed tip cells, whereas *Sirt1* mutant cells, exhibiting increased NICD levels, preferentially formed stalk cells (Fig. 4i–m). Furthermore, inhibition of Sirt1 by NAM in *Dll4* heterozygote EBs (*Dll4*^{+/lacZ}), which are characterized by reduced Notch signalling and excessive vessel branching¹⁹, restored normal branching behaviour (Supplementary Fig. 13), indicating that SIRT1 inhibition can re-adjust adequate levels of Notch signalling in *Dll4*^{+/lacZ} cells.

Our studies unravel a novel role for SIRT1 as a cell-autonomous negative modulator of Notch signalling and highlight reversible acetylation of NICD as a key molecular mechanism to adjust the dynamics of Notch responses. Negative regulation of Notch signalling by SIRT1 might, however, involve additional components of the Notch pathway such as the co-repressor NCOR and transcription factors of the HES/HEY and FOXO family, which have been shown to be regulated by SIRT1 (refs 20–24). By cooperative regulation of these factors that function at different steps in the Notch pathway, SIRT1 might synergistically modulate Notch signalling to achieve robust regulation of Notch-controlled cellular responses. Notably, the regulatory function of SIRT1 in the Notch pathway seems to extend beyond NOTCH1 and to include other family members such as NOTCH4 (Supplementary Fig. 14), indicating a general mechanism for Notch regulation by SIRT1 that may apply to several Notch-controlled biological processes. However, we cannot rule out that SIRT1 might also indirectly affect Notch signalling in other tissues or cellular environments, as shown in ageing neuronal cells, where SIRT1 has been found to increase Notch signalling through a RXR-dependent regulation of ADAM10 (ref. 25). However, in endothelial cells *ADAM10* expression was not affected by SIRT1 (Supplementary Fig. 15), indicating that the indirect modulation of Notch signalling through RXR is not operational in endothelial cells. Because SIRT1 requires NAD⁺ for its catalytic activity, it is responsive to changes in the metabolic and redox state of the cell²⁶ and might therefore function as a cellular sensor coupling energy and oxygen homeostasis to Notch-dependent control of branching vessel morphogenesis.

METHODS SUMMARY

Human umbilical vein endothelial cells (HUVECs) were from Lonza and cultured as described⁸. Mouse lung endothelial cells were isolated, purified and cultured as described⁸. The culture of ES cells and the generation of EBs were performed as described¹⁸. Endothelial-restricted *Sirt1* mutant mice were on a C57BL/6 genetic background and generated as described⁸. *Tg(fli1a:nEGFP)y7* and *Tg(fli1a:EGFP)y1* zebrafish lines were used, maintained and bred under standard conditions. Antisense morpholino oligonucleotides targeting *sirt1* and *dll4* were used as described^{9,19}. Mass spectrometric experiments were performed on a nano-flow HPLC system (Proxeon) connected to a LTQ-Orbitrap Velos instrument (Thermo Fisher Scientific) equipped with a nano-electrospray source (Proxeon).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 October 2009; accepted 10 February 2011.

Published online 18 April; corrected 20 April 2011 (see full-text HTML version for details).

1. Kopan, R. & Ilagan, M. X. The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell* **137**, 216–233 (2009).
2. Roca, C. & Adams, R. H. Regulation of vascular morphogenesis by Notch signaling. *Genes Dev.* **21**, 2511–2524 (2007).
3. Phng, L. K. & Gerhardt, H. Angiogenesis: a team effort coordinated by notch. *Dev. Cell* **16**, 196–208 (2009).
4. Milne, J. C. *et al.* Small molecule activators of SIRT1 as therapeutics for the treatment of type 2 diabetes. *Nature* **450**, 712–716 (2007).

5. Kurooka, H. & Honjo, T. Functional interaction between the mouse Notch1 intracellular region and histone acetyltransferases PCAF and GCN5. *J. Biol. Chem.* **275**, 17211–17220 (2000).
6. Oswald, F. *et al.* p300 acts as a transcriptional coactivator for mammalian Notch-1. *Mol. Cell. Biol.* **21**, 7761–7774 (2001).
7. Kim, M. Y. *et al.* Tip60 histone acetyltransferase acts as a negative regulator of Notch1 signaling by means of acetylation. *Mol. Cell. Biol.* **27**, 6506–6519 (2007).
8. Potente, M. *et al.* SIRT1 controls endothelial angiogenic functions during vascular growth. *Genes Dev.* **21**, 2644–2658 (2007).
9. Hellström, M. *et al.* Dll4 signalling through Notch1 regulates formation of tip cells during angiogenesis. *Nature* **445**, 776–780 (2007).
10. Ridgway, J. *et al.* Inhibition of Dll4 signalling inhibits tumour growth by deregulating angiogenesis. *Nature* **444**, 1083–1087 (2006).
11. Suchting, S. *et al.* The Notch ligand Delta-like 4 negatively regulates endothelial tip cell formation and vessel branching. *Proc. Natl Acad. Sci. USA* **104**, 3225–3230 (2007).
12. Lobov, I. B. *et al.* Delta-like ligand 4 (Dll4) is induced by VEGF as a negative regulator of angiogenic sprouting. *Proc. Natl Acad. Sci. USA* **104**, 3219–3224 (2007).
13. Phng, L. K. *et al.* Nrarp coordinates endothelial Notch and Wnt signaling to control vessel density in angiogenesis. *Dev. Cell* **16**, 70–82 (2009).
14. Benedetto, R. *et al.* The notch ligands Dll4 and Jagged1 have opposing effects on angiogenesis. *Cell* **137**, 1124–1135 (2009).
15. Bentley, K., Mariggi, G., Gerhardt, H. & Bates, P. A. Tipping the balance: robustness of tip cell selection, migration and fusion in angiogenesis. *PLoS Comput. Biol.* **5**, e1000549 (2009).
16. Siekmann, A. F. & Lawson, N. D. Notch signalling limits angiogenic cell behaviour in developing zebrafish arteries. *Nature* **445**, 781–784 (2007).
17. Leslie, J. D. *et al.* Endothelial signalling by the Notch ligand Delta-like 4 restricts angiogenesis. *Development* **134**, 839–844 (2007).
18. Jakobsson, L. *et al.* Heparan sulfate in trans potentiates VEGFR-mediated angiogenesis. *Dev. Cell* **10**, 625–634 (2006).
19. Jakobsson, L. *et al.* Endothelial cells dynamically compete for the tip cell position during angiogenic sprouting. *Nature Cell Biol.* **12**, 943–953 (2010).
20. Picard, F. *et al.* Sirt1 promotes fat mobilization in white adipocytes by repressing PPAR-gamma. *Nature* **429**, 771–776 (2004).
21. Hisahara, S. *et al.* Histone deacetylase SIRT1 modulates neuronal differentiation by its nuclear translocation. *Proc. Natl Acad. Sci. USA* **105**, 15599–15604 (2008).
22. Takata, T. & Ishikawa, F. Human Sir2-related protein SIRT1 associates with the bHLH repressors HES1 and HEY2 and is involved in HES1- and HEY2-mediated transcriptional repression. *Biochem. Biophys. Res. Commun.* **301**, 250–257 (2003).
23. Prozorovski, T. *et al.* Sirt1 contributes critically to the redox-dependent fate of neural progenitors. *Nature Cell Biol.* **10**, 385–394 (2008).
24. Kitamura, T. *et al.* A Foxo/Notch pathway controls myogenic differentiation and fiber type specification. *J. Clin. Invest.* **117**, 2477–2485 (2007).
25. Donmez, G., Wang, D., Cohen, D. E. & Guarente, L. SIRT1 suppresses β -amyloid production by activating the alpha-secretase gene ADAM10. *Cell* **142**, 320–332 (2010).
26. Finkel, T., Deng, C. X. & Mostoslavsky, R. Recent progress in the biology and physiology of sirtuins. *Nature* **460**, 587–591 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are thankful to F. W. Alt, R. Kopan, Z. Lou, E. Seto, S. L. Berger, S. Diane Hayward, S. McMahon, G. Thurston and N. D. Lawson for reagents and to I. Dikic for comments. This work was supported by grants from the DFG (PO1306/1-1, SFB 834/A6 and Exc 147/1). F.D. was supported by the Interuniversity Attraction Poles Program–Belgian Science Policy (IUAP-BELSP0 PVI/28). R.M. is supported by the Sidney Kimmel Cancer Research Foundation, a New Investigator Grant from the Massachusetts Life Sciences Center, an AFAR Research Grant and NIH grants (R01DK088190-01A1 and R01GM093072-01). H.G. is supported by Cancer Research UK, the European Molecular Biology Organisation Young Investigator Programme, and The Lister Institute of Preventive Medicine. H.G. and K.B. are supported by the Fondation Leducq Transatlantic Network of Excellence ARTEMIS. C.A.F. is supported by the Marie Curie FP7 People initiative. G.D. and M.M. thank F. Pezzimenti for fish care and technical help, and AIRC (Associazione Italiana per la Ricerca sul Cancro) for financial support.

Author Contributions V.G. and M.P. designed and guided research. V.G., G.D., C.A.F., M.K., L.-K.P., K.B., L.T., F.D., M.H.S., B.Z., R.P.B., M.M., H.G. and M.P. performed research. V.G., G.D., C.A.F., M.K., L.-K.P., K.B., L.T., F.D., M.M., H.G. and M.P. analysed data. R.M., C.H.W. and T.B. provided reagents and/or technical support. T.B., A.M.Z. and S.D. gave conceptual advice. V.G., H.G. and M.P. wrote the paper. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.P. (potente@em.uni-frankfurt.de).

METHODS

Cells and cell culture. Pooled human umbilical vein endothelial cells (HUVECs) were purchased from Lonza and cultured as described²⁷. HEK293T cells were purchased from ATCC and Invitrogen and cultured as recommended. Mouse lung endothelial cells (MLECs) were isolated, purified and cultured as described⁴⁸.

DLL4 stimulation of endothelial cells. Lyophilized recombinant human or mouse DLL4 was purchased from R&D Systems and reconstituted at $100 \mu\text{g ml}^{-1}$ in PBS containing 0.1% bovine serum albumin. For stimulation of cultured endothelial cells, DLL4 was immobilized by coating culture dishes with $500 \mu\text{g ml}^{-1}$ DLL4 in PBS for 1 h at room temperature or overnight at 4°C .

Reagents and pharmacological *in vitro* cell treatments. HUVECs or HEK293 cells were pharmacologically treated with $10 \mu\text{M}$ trichostatin A (TSA; Calbiochem), 20 mM nicotinamide (NAM; Sigma), $10 \mu\text{M}$ MG132 (Calbiochem) or $50 \mu\text{g ml}^{-1}$ cycloheximide (CHX; Calbiochem). SRT2183 was used at a concentration of $20 \mu\text{M}$ (Sirtis Pharmaceuticals). Control groups were treated with the respective vehicles. To inhibit Notch signalling *in vitro*, HUVECs were incubated with $0.08 \mu\text{M}$ DBZ ((S,S)-2-[2-(3,5-Difluorophenyl)acetylaminol]-N-(5-methyl-6-oxo-6,7-dihydro-5H-dibenzo[b,d]azepin-7-yl)propionamide). When HUVECs were co-stimulated with DLL4, cells were pre-treated for at least 1 h before being replated on DLL4-coated dishes.

Plasmids and transfections. The intracellular domain of murine Notch1 (bp 5,230 to bp 7,593) and Notch4 (bp 4,417 to bp 5,895) was cloned in-frame into the mammalian expression vectors pcDNA3.1/nV5-DEST or pDEST40 as well as pcDNA3.0 N-Flag and pDEST490 harbouring V5 or Flag tags at the N or C terminus, respectively. Human SIRT1, SIRT2, SIRT3 and SIRT5 were subcloned in a derivative of the pcDNA3.1(+) vector to generate C-terminal Flag-tagged fusions as described²⁸. The serial deletions mutants of SIRT1 were provided by Z. Lou²⁹, the NICD(ΔC)-Myc by R. Kopan, the p300 expression plasmid by R. Eckner, PCAF-Flag by S. L. Berger and TIP60-Flag by S. McMahon. The Notch-regulated luciferase reporter genes TP1 and 4×CBF1 were from U. Zimmer-Strobl and D. Hayward, respectively. Transient transfections of HEK293 cells were carried out with Lipofectamine 2000 transfection reagent (Invitrogen). HUVECs were grown to 60–70% confluence and transfected with the Trans Pass V reagents (New England Biolabs) as recommended.

RNA interference. To silence *SIRT1* gene expression, cells were transfected with a specific *SIRT1* siRNA synthesized by Eurogentec (5'-GAAGTTGACCTCCTC ATTG-3') or a validated pool of siRNA duplexes directed against human *SIRT1* (On-Target plus SMART pool), which was purchased from Dharmacon. A scrambled siRNA was used as a control (5'-TTCTCCGAACGTGCGACGA-3'). HUVECs and HEK293 cells were transfected with the indicated siRNAs (50 nM) using the GeneTrans II reagent (MoBiTec) or Lipofectamine 2000 (Invitrogen), respectively.

Luciferase assays. Reporter assays in HUVECs were performed with the Dual-Luciferase Reporter Assay System (Promega) and a LUMAT LB 9507 luminometer (BERTHOLD Technologies). Briefly, 24 h after co-transfection with the Notch luciferase reporters and the constitutive Renilla luciferase reporter pGL4.74hRluc/TK (Promega) HUVECs were lysed and reporter assays performed according to the manufacturers' protocols. For siRNA experiments, cells were transfected with siRNAs and after 24 h transfected with the Notch luciferase reporters and the constitutive Renilla luciferase reporter. For experiments in which Notch activity was induced by DLL4, transfected HUVECs were replated on DLL4-coated dishes 6 h after plasmid transfections. Luciferase activity was measured after an additional 24 h. Reporter activity was adjusted for the internal Renilla luciferase control and is expressed relative to control.

RNA analysis by real-time quantitative PCR. RNA from HUVECs or MLECs was isolated with the RNeasy Kit (Qiagen) as recommended. TaqMan Gene Expression Assays for *HEY2/Hey2* (Hs00232622_m1; Mm00469280_m1), *NRARP/Nrarp* (Hs01104102_m1; Mm00482529_m1), *Lfng* (Mm00456128_m1), *NOTCH1* (Hs01062014_m1) and *ADAM10/Adam10* (Hs00153853_m1; Mm00545742_m1) were obtained from Applied Biosystems and qPCR was carried out using the StepOnePlus real-time PCR system (Applied Biosystems).

Co-immunoprecipitations. For co-immunoprecipitations of overexpressed proteins, HEK293 cells were lysed 24 h post-transfection in IPLS buffer (50 mM Tris-HCl pH7.5, 120 mM NaCl, 0.5 mM EDTA and 0.5% Nonidet P-40). After pre-clearing, immunoprecipitations were performed using direct-conjugated immuno-affinity agarose beads against Flag (Sigma), Myc (Sigma) or V5 (Invitrogen) at 4°C with gentle rotation over night. After western blotting, the presence of the co-immunoprecipitated proteins was analysed with antibodies against Flag (1:1000, Sigma), Myc (1:1000, Santa Cruz) and V5 (1:5000, Invitrogen). For co-immunoprecipitation of endogenous SIRT1 with endogenous NICD, HUVECs were pretreated with $10 \mu\text{M}$ MG132 (Calbiochem) for 1 h and then re-plated on DLL4-coated culture dishes in the presence of MG132 for 6 h. Cells were lysed in IPLS buffer. After pre-clearing, SIRT1 was immunoprecipitated with a mouse

anti-human SIRT1 monoclonal antibody (Santa Cruz) for 12 h at 4°C followed by incubation with A/G agarose beads (Santa Cruz) for an additional 2 h. Immune complexes were analysed by western blotting using NICD (1:1,000, Cell Signaling Technology) and SIRT1 (1:1,000, Cell Signaling Technology) antibodies.

Detection of NICD acetylation. For detection of acetylated NICD in HEK293 cells, NICD-V5-transfected HEK293 cells were treated with NAM, TSA, the respective vehicles or combinations thereof for 6 or 16 h. Acetylation was also measured in *SIRT1* siRNA transfected cells 24 h after NICD-V5 overexpression. For both approaches, cells were lysed in RIPA buffer supplemented with $10 \mu\text{M}$ TSA and 20 mM NAM (AcRIPA). After pre-clearing cell extracts were subjected to immunoprecipitation using agarose beads coupled to a V5 antibody. The immune complexes were analysed by western blotting with polyclonal anti-acetylated lysine antibodies (AcK, 1:300, Cell Signaling or 1:1,000, Abcam). For detection of endogenous acetylated NICD, HUVECs were plated on DLL4-coated culture dishes and stimulated for 6 h in the presence or absence of NAM. HUVECs were lysed in AcRIPA buffer. After pre-clearing, NICD was purified from total cell lysates with a Notch1-specific antibody (Cell Signaling) for 12 h at 4°C followed by incubation with protein A/G agarose beads (Santa Cruz) for an additional 2 h. Western blotting was performed with an antibody targeting acetylated lysines (1:300, Cell Signaling).

***In vitro* deacetylation assay.** NICD(ΔC)-Myc was expressed in HEK293 cells alone or together with p300 and immunopurified using an anti-Myc antibody. Immunoprecipitated NICD(ΔC)-Myc was four times in IPLS buffer and once in ST buffer (50 mM Tris-HCl, pH9, 4 mM MgCl_2 and 0.2 mM DTT). Purified NICD(ΔC)-Myc was then incubated with recombinant wild-type or deacetylase defective (H363Y) SIRT1 in $30 \mu\text{l}$ of ST buffer containing 800 nM TSA and supplemented or not with $50 \mu\text{M}$ of NAD^+ . Reactions were incubated at 30°C under robust agitation for 2 h and stopped by addition of Leammli buffer. Acetylation of NICD was then assessed by western blotting using an anti-acetylated lysine antibody.

Ubiquitination assay. For detection of ubiquitinated NICD in HEK293 cells, NICD-V5 was overexpressed in HEK293 cells together with HA-tagged ubiquitin. The effect of SIRT1 on NICD ubiquitination was assessed after knockdown of *SIRT1* by siRNA or after NAM treatment. Cells were lysed in RIPA buffer and subjected to immunoprecipitation using anti-V5-coupled agarose beads. The immune complexes were analysed by western blotting with anti-HA antibodies (HA, 1:1000, Roche). For quantification of NICD ubiquitination the anti-HA signal was normalized to the NICD-V5 levels.

Western blot analysis and antibodies. SDS-PAGE and western blot analyses were performed according to standard procedures and detected with the ECL detection kit (GE Health Care Bio-Sciences). Quantification of band intensities by densitometry was carried out using the Image J software.

Mass spectrometric analysis. HEK293 cells were transfected with NICD-V5 and PCAF-Flag and 24 h after transfection, treated with nicotinamide for 6 h. After NICD-V5 immunoprecipitation proteins were separated by one-dimensional SDS-PAGE (4–12% Novex-gels, Invitrogen) and stained with colloidal Coomassie. NICD-V5 gel bands were excised and subjected to in-gel digest with trypsin³⁰. The resulting tryptic peptides were extracted with acetonitrile, and desalted with reversed phase C18 STAGE tips³¹. Mass spectrometric experiments were performed on a nano-flow HPLC system (Proxeon) connected to a LTQ-Orbitrap Velos instrument (Thermo Fisher Scientific) equipped with a nanoelectrospray source (Proxeon). The mass spectrometer was operated in the data dependent mode to monitor MS and MS/MS spectra. Survey full-scan MS spectra (from m/z 300–2,000) were acquired in the Orbitrap with a resolution of $R = 60,000$ at m/z 400 after accumulation of 1,000,000 ions. The five most intense ions from the preview survey scan delivered by the Orbitrap were sequenced by collision-induced dissociation (CID) in the LTQ. For higher C-trap dissociation (HCD) 30,000 ions were accumulated in the C-trap and MS/MS spectra were detected in the orbitrap at a resolution of 7,500 (ref. 32). Mass spectra were analysed using MaxQuant software³³ and automated database searching (Matrix Science). All tandem mass spectra were searched against the mouse International Protein Index protein sequence database (IPI version 3.54) and concatenated with reversed copies of all sequences. The required false positive rate was set to 1% at the protein level, and maximum allowed mass deviation was set to 5 p.p.m. in MS mode and 0.5 Da for MS/MS peaks. Cysteine carbamidomethylation was searched as a fixed modification and N-acetyl protein, oxidized methionine and acetylation of lysine was searched as variable modifications. A maximum of three missed cleavages was allowed.

Site-directed mutagenesis. The acetylation-defective NICD(14KR)-V5 mutant was generated by site-directed mutagenesis replacing the acetylated lysines by arginine using the QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent technologies). Primer sequences are available upon request. In addition, we generated a synthetic gene encoding for the murine NICD sequence harbouring the lysine-to-arginine substitutions at the sites identified by mass spectrometry

(Geneart). Using Gateway cloning the synthesized gene was cloned in frame into the mammalian expression vector pDEST490 containing a V5 tag at the C terminus.

Endothelial cell proliferation and three-dimensional sprouting assay. Cell growth was assayed by a colorimetric procedure with the cell counting kit-8 (Dojindo Laboratories), modified from a method that uses 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide as recommended. HUVECs were cultured in DLL4 or solvent and treated with DBZ (0.08 μ M) or DMSO (solvent). The cell counting kit-8 reagent was added to the cultures according to manufacturer's instructions and growth assessed after 0, 24, 48, 72 and 96 h in culture. Endothelial cell spheroids of defined cell number were generated as described previously²⁷ and sprouting assessed after 24 h.

Mouse embryonic stem cells and embryoid bodies assay. DsRed-MST ES cells and *Dlla4^{+/-lacZ}* ES cells were obtained from A. Nagy and G. Thurston (Regeneron Pharmaceuticals), respectively, and have been described previously^{34,39}. The generation of *Sirt1* mutant ES cells was previously described³⁵. The lack of exon 4 in the *Sirt1* gene generates a truncated protein that is catalytically inactive and both cells and mice exhibit the same phenotype as the full *Sirt1* knockout, as previously described³⁵. ES cell culture and generation of EBs were performed as described¹⁸. The following primary antibodies were used for detection: rat anti-mouse CD31 antibody (BD Biosciences), rabbit anti-mouse cleaved Notch1 (Abcam). Secondary antibodies: Alexa 647 donkey anti-rat IgG, Alexa 488 donkey anti-rabbit, (Molecular Probes), and DAPI (Sigma) for staining of nuclei.

Quantification analysis of tip cells and NICD levels in EB sprouts. Complete high-resolution three-dimensional rendering of EBs was acquired using a Laser Scanning Microscope 710 (Zeiss) equipped with a motorized stage. For quantification of tip cells, 50–200 tip cells per EB were manually scored and marked for genotypic origin using the Imaris 7.0.0 software. Tip regions were defined as the regions from the very tips to the second or third vascular loop. Quantification was done using Prism 5.0c software. For the quantification of NICD protein levels, complete 1,024 \times 1,024, 12 bit images of vascular sprouts were collected with \times 63 objective lens using a Laser Scanning Microscope 710 (Zeiss), keeping the same settings for all images (Supplementary Fig. 16). Images were processed using the Imaris 7.0.0 software. To identify nuclear NICD spots we used 'Surface' Imaris function. Measurements of number of spots, volume and total pixel intensity of each spot were taken and statistics were analysed using Prism 5.0c software.

Zebrafish strains, maintenance and drug treatment. Zebrafish strains were maintained and bred under standard conditions. AB wild type strain, *Tg(fli1a:nEGFP)y7* and *Tg(fli1a:EGFP)y1*³⁶ lines were used. Fish were raised/maintained according to EU regulations on laboratory animals. DAPT (Sigma) was dissolved in DMSO and added to E3 media to give a final concentration of 100 μ M DAPT plus 0.2% DMSO. Embryos were dechorionated and incubated at 28.5 °C in E3 containing 100 μ M DAPT and 0.2% DMSO from the end of gastrulation until 28 h.p.f. E3 containing DMSO alone was used as a control.

Antisense morpholino oligonucleotides, fluorescence microscopy and cell counts. Antisense morpholino oligonucleotides against the exon1–intron1 splice site, 5'-CGAACCAAACTACCAATCTGTGGC-3', specifically targeting zygotic *sirt1* (MO2-*sirt1*, a splice blocking MO)⁸, and the 5'-GTTCGAGCTTACCGG CCACCAAG-3', specifically targeting zygotic *dlla4* (MO1-*dlla4*, a splice blocking MO)¹⁶ were used. For cell counting experiments, 28 h.p.f. *Tg(fli1a:nEGFP)y7* embryos were fixed in 1% paraformaldehyde in PBS containing CaCl₂ 0.15 mM and 4% sucrose overnight at 4 °C and were then washed in PBS at room temperature. Embryos were permeabilized for 1–3 h at room temperature and then mounted in 85% glycerol solution. Using confocal microscope, endothelial cells of *Tg(fli1a:nEGFP)y7* were counted from projections of Z-series of optical sections. The region counted in each specimen spanned two somites (three intersegmental vessels).

Immunofluorescence and analysis of postnatal angiogenesis in the mouse retina. Endothelial-restricted *Sirt1* mutant mice were on a C57BL/6 genetic background and generated as described⁸. Mice were killed at postnatal day 5 (P5) and their eyes collected for analysis. To inhibit Notch signalling in mice, DAPT (Calbiochem) was dissolved in 10% ethanol and 90% corn oil. The DAPT solution was then injected intraperitoneally at a dose of 100 mg kg⁻¹ per mouse. Control mice were injected with the same amount of vehicle. Mice were treated for 48 h. Injections were performed once a day, at P3 and P4 and eyes harvested at P5. For

whole-mount analysis of retinal blood vessels, retinas were fixed in 4% PFA for 2 h at 4 °C, or in MeOH at -20 °C. Antibodies were diluted in PBS containing 0.5% BSA and 0.25% Tween 20 except for isolectin-B4, which was diluted in PBlec (PBS (pH 6.8), 1% Triton X-100, 1 mM CaCl₂, 1 mM MgCl₂, 1 mM MnCl₂). Primary antibodies used: biotinylated-isolectin-B4 (Vector Laboratories and Molecular Probes), BrdU (Molecular Probes) and SIRT1 (Upstate). Alexa-488- or 568-conjugated secondary antibodies were used (Molecular Probes). For labelling of proliferating cells, 50 mg kg⁻¹ of BrdU (BD Biosciences) per pup was injected intraperitoneally 3 h before they were killed. Retinas were fixed for 2 h in 4% PFA and then incubated for 2 h in 2N HCl. Retinas were blocked with 1% BSA/0.5% Tween 20 in PBS and incubated overnight at 4 °C with a mouse monoclonal anti-BrdU antibody. Secondary detection was performed with goat anti-mouse Alexa Fluor-coupled secondary antibody. After BrdU staining, retinas were fixed for 30 min in 4% PFA, washed and then stained with directly conjugated Isolectin-B4. Retinas were flat mounted (Dako fluorescent mounting medium) and confocal laser scanning microscopy was performed using Zeiss LSM 510 microscope. All quantifications were done with high-resolution confocal images using Z-sections of the sample. For quantifying vessel migration, the distance of vessel growth from the optic nerve to the periphery was measured using the ImageJ software. For vessel density, the number of vessel branch points was counted from five fields per retina sized 100 \times 100 μ m, from at least 10 retina samples per group. The number of filopodial extensions was quantified at the angiogenic front and counted in at least 18 fields (sized 132 \times 132 μ m, 6 retinas per group). The total number of filopodia was normalized to 100 μ m of vessel length at the angiogenic front. BrdU-labelled isolectin B4-positive endothelial cells were counted in 30 fields (sized 450 \times 450 μ m, 6 retinas per group).

Computational modelling of endothelial tip-cell selection. A vessel segment comprising seven cells, with one cell per vessel cross-section, was used with all parameters set as described¹⁵. *Sirt1* loss was modelled by extending the number of time steps that the 'current active Notch' level, which represents NICD in the model, affected downstream targets. In the selection model, Notch activity leads to downregulation of VEGFR2 receptors. It was only effective for one time step when modelling wild-type vessels (one time step representing 15 s in the model). To match the observed 2.5-fold increase in NICD due to the loss of *Sirt1* (referred to as *Sirt1* null vessels), the time the current active Notch level was effective for was extended to three time steps; however, on the third time step only half of that NICD level continued to affect VEGFR2 downregulation. To investigate the sensitivity and effect of increasing the NICD lifetime, further simulations were performed where it was set to a hypothetical 4.5- and then 9.5-fold increase, using a similar approach.

Statistical analysis. Statistical analysis was performed by ANOVA followed by Bonferroni's multiple comparison test, or Student's *t*-test using Prism 5 (GraphPAD Software Inc.). *P* values of <0.05 were considered significant.

27. Potente, M. *et al.* Involvement of Foxo transcription factors in angiogenesis and postnatal neovascularization. *J. Clin. Invest.* **115**, 2382–2392 (2005).
28. North, B. J. *et al.* The human Sir2 ortholog, SIRT2, is an NAD⁺-dependent tubulin deacetylase. *Mol. Cell* **11**, 437–444 (2003).
29. Kim, J. E., Chen, J. & Lou, Z. DBC1 is a negative regulator of SIRT1. *Nature* **451**, 583–586 (2008).
30. Shevchenko, A. *et al.* In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols* **1**, 2856–2860 (2006).
31. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols* **2**, 1896–1906 (2007).
32. Olsen, J. V. *et al.* A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* **8**, 2759–2769 (2009).
33. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.* **26**, 1367–1372 (2008).
34. Vintersten, K. *et al.* Mouse in red: red fluorescent protein expression in mouse ES cells, embryos, and adult animals. *Genesis* **40**, 241–246 (2004).
35. Cheng, H. L. *et al.* Developmental defects and p53 hyperacetylation in *Sirt1* homolog (*SIRT1*)-deficient mice. *Proc. Natl Acad. Sci. USA* **100**, 10794–10799 (2003).
36. Lawson, N. D. & Weinstein, B. M. *In vivo* imaging of embryonic vascular development using transgenic zebrafish. *Dev. Biol.* **248**, 307–318 (2002).

Preserving the membrane barrier for small molecules during bacterial protein translocation

Eunyong Park¹ & Tom A. Rapoport¹

Many proteins are translocated through the SecY channel in bacteria and archaea and through the related Sec61 channel in eukaryotes¹. The channel has an hourglass shape with a narrow constriction approximately halfway across the membrane, formed by a pore ring of amino acids². While the cytoplasmic cavity of the channel is empty, the extracellular cavity is filled with a short helix called the plug², which moves out of the way during protein translocation^{3,4}. The mechanism by which the channel transports large polypeptides and yet prevents the passage of small molecules, such as ions or metabolites, has been controversial^{2,5–8}. Here, we have addressed this issue in intact *Escherichia coli* cells by testing the permeation of small molecules through wild-type and mutant SecY channels, which are either in the resting state or contain a defined translocating polypeptide chain. We show that in the resting state, the channel is sealed by both the pore ring and the plug domain. During translocation, the pore ring forms a 'gasket-like' seal around the polypeptide chain, preventing the permeation of small molecules. The structural conservation of the channel in all organisms indicates that this may be a universal mechanism by which the membrane barrier is maintained during protein translocation.

Bacteria offer a unique opportunity to test the permeation of small molecules through the protein translocation channel because the channel is located in the plasma membrane and is therefore accessible in intact cells. To test the permeability of the resting SecY channel, we compared *E. coli* wild-type SecY, which is expected to be sealed, with a plug-deletion mutant of SecY (SecY(Δ P)), which should be constitutively open (Supplementary Fig. 1). Although a new plug may form from neighbouring polypeptide segments in this mutant⁹, it probably blocks the channel only transiently⁸. Wild-type and SecY(Δ P) channels were expressed under an inducible promoter at about the same level as the endogenous protein and expression of the SecY(Δ P) mutant caused only a moderate growth defect (Supplementary Fig. 2).

First, we studied the permeation of a relatively large (525 Da), uncharged cysteine-modifying reagent, biotin-PEG₂-maleimide (BM; Fig. 1a), which can cross the outer membrane through porins but cannot cross the inner membrane¹⁰ (Supplementary Fig. 1). When BM was added to wild-type *E. coli* cells, few proteins were biotinylated (Fig. 1b, lane 5). By contrast, in the SecY(Δ P) mutant, several proteins were strongly modified, particularly a protein of about 30 kDa (Fig. 1b, lane 8). Most of the modified proteins were located in the cytosol, as demonstrated by cell fractionation (Fig. 1c). The extent of modification was about the same after treatment with the transcription inhibitor rifampicin (Fig. 1b, lane 9), which clears all SecY channels of translocating polypeptides (see Fig. 2c). Thus, permeation of BM occurs primarily through resting SecY(Δ P) channels. We also found that many signal sequence suppressor (*prl*) SecY mutants allow the permeation of BM, although to a lesser extent than the SecY(Δ P) mutant (Supplementary Fig. 3). Channel opening in the absence of a translocation substrate explains why these mutants translocate proteins with defective or missing signal sequences^{11–13}.

Next, we used osmotic swelling and bursting of cells (Supplementary Fig. 1) to test the permeability of the resting SecY channel to

xylitol, an uncharged sugar of 152 Da. *Escherichia coli* cells were converted to spheroplasts and diluted into an iso-osmotic solution of xylitol. Spheroplasts containing wild-type SecY channels did not take up xylitol and therefore the turbidity of the sample changed little over time (Fig. 1d). In contrast, the SecY(Δ P) mutant allowed xylitol to permeate, particularly when the channel was cleared of translocating chains by rifampicin (Fig. 1d). Finally, we used osmotic swelling and bursting to test the permeation of charged and small (35 Da) Cl[−] ions. Spheroplasts were diluted into an iso-osmotic solution of KCl in the presence of valinomycin, an ionophore that allows the K⁺ counterions to move directly through the lipid bilayer. The data showed that wild-type SecY did not conduct Cl[−] ions, in contrast to the SecY(Δ P) mutant (Fig. 1e). We conclude that the resting wild-type channel is impermeable to the small molecules tested and that the plug domain of

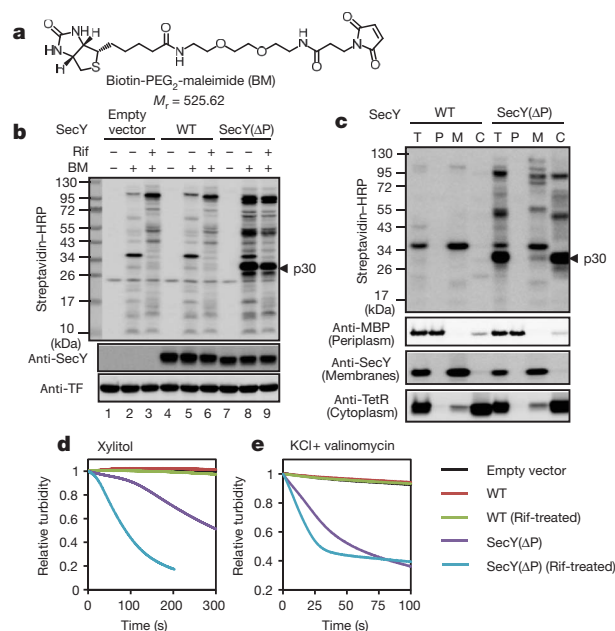


Figure 1 | Testing the permeability of the resting SecY channel. **a**, Structure of the modification reagent BM. **b**, Wild-type (WT) SecY or the SecY(Δ P) mutant were expressed under the inducible Tet promoter. Cells were incubated with BM and the proteins were separated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE), then blotted with streptavidin–HRP conjugate, SecY antibodies or trigger factor (TF) antibodies (loading control). Where indicated, rifampicin (Rif) was added before BM. Endogenous SecY was tagged at its C terminus, abolishing recognition by SecY antibodies (Supplementary Fig. 2). p30, a prominent biotinylated protein. **c**, Experiment conducted as in **b**, but cells (T) were fractionated into periplasm (P), membranes (M) and cytosol (C) after incubation with BM. Fractionation was controlled by immunoblotting for the indicated marker proteins. MBP, maltose-binding protein; TetR, tetracycline repressor. **d**, **e**, Spheroplasts were diluted into iso-osmotic solutions of xylitol (**d**) or KCl containing valinomycin (**e**) and the change in turbidity was followed over time.

¹Howard Hughes Medical Institute and Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA.

SecY is required for the seal. It should be noted that the SecY(Δ P) mutant did not allow passage of K^+ (Supplementary Fig. 4), Na^+ or SO_4^{2-} ions (not shown). Thus, in agreement with previous results¹⁴, the open channel still provides a barrier to some molecules (see Supplementary Discussion), which may explain the relatively minor growth defect of the SecY(Δ P) mutant (Supplementary Fig. 2).

To study the permeability of the active SecY channel, we developed a method to occupy the channels *in vivo* with a defined co-translational translocation intermediate. The model substrate (NC100) contains 100 amino acids (Fig. 2a) including the signal sequence of DsbA, which targets it to the signal recognition particle (SRP)-dependent co-translational translocation pathway¹⁵. NC100 also contains a sequence from an unrelated protein, a Myc tag for detection and the SecM-stalling sequence^{16–18}. After synthesis of the SecM sequence, the ribosome stalls on the mRNA with the nascent chain associated as peptidyl-tRNA (Fig. 2a). This construct was synthesized from an inducible promoter in cells expressing the SecY channel from a constitutive promoter. The insertion of the nascent chain into the channel was verified by *in vivo* disulphide crosslinking: addition of an oxidant to the cell culture led to efficient crosslinking of a single cysteine at position 19 in NC100 to a single cysteine at position 68 of the plug domain of SecY (SecY_{68C}; Supplementary Fig. 5). Channel insertion was dependent on the hydrophobicity of the signal sequence and was strongly reduced in the absence of SRP (Supplementary Fig. 6). When the expression of SecY_{68C} was diminished by changing the start codon, about 70% of SecY_{68C} was crosslinked to NC100, as judged by the

reduction of the non-crosslinked SecY band upon addition of the oxidant (Fig. 2b, lane 2 versus lane 1; quantification was confirmed by loading different amounts, see lanes 7–9). When NC100 expression was not induced, a lower percentage of SecY_{68C} was crosslinked to endogenous proteins (Fig. 2b, lane 4 versus lane 2; white arrowheads); these crosslinks disappeared over time in the presence of rifampicin (Fig. 2b, lanes 6 and 11). These results indicate that most of the SecY molecules can be occupied by NC100. Given the almost 1:1 molar ratio of nascent chain to SecY, a single SecY copy may be sufficient for co-translational translocation of a nascent chain, as was previously proposed^{19,20}.

We used the new method to ask whether the open SecY channel, represented by the SecY(Δ P) mutant (see Fig. 1), can be blocked for small molecules by NC100. The SecY(Δ P) mutant expressed from a constitutive promoter was leaky for the modification reagent BM but induction of NC100 abolished permeation (Fig. 2c, lane 8 versus lane 7). Inhibition of transcription by rifampicin released NC100 from ribosomes and restored leakiness for BM (lane 9). NC100 with a defective signal sequence containing two arginines in the hydrophobic core (RR) did not block BM permeation (lane 10). With the wild-type SecY channel, no leakage was observed regardless of whether or not NC100 was expressed (lanes 1–5). Thus, the open pore of the SecY(Δ P) mutant is sealed upon binding of the ribosome–nascent-chain complex to the SecY channel.

To test whether the seal is provided by channel insertion of the nascent polypeptide, we expressed chains of different lengths. Nascent

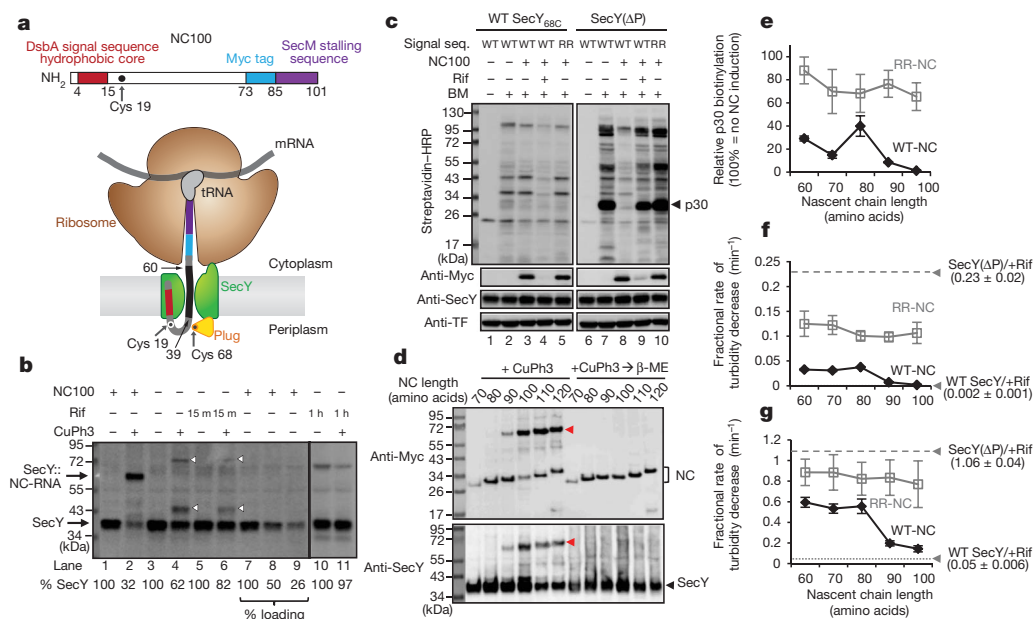


Figure 2 | Testing the permeability of a translocating SecY channel.

a, Schematics of the model nascent chain NC100 (upper panel) and its insertion into the SecY channel (lower panel). The cysteines indicated in NC100 and SecY form a disulphide bridge. Residues 39–60 are inside the channel. **b**, NC100 with Cys 19 was expressed under an arabinose-inducible promoter together with SecY_{68C} under the endogenous promoter. The start codon of SecY_{68C} was changed from AUG to GUG. Where indicated, cells were treated with the oxidant copper(II)(1, 10-phenanthroline)₃ (CuPh₃). Rifampicin was added before oxidation for 15 min or 1 h, as indicated. The samples were analysed by SDS-PAGE, followed by blotting with antibodies against SecY::NC-RNA, SecY crosslinked to tRNA-associated NC100. White arrowheads indicate SecY_{68C} crosslinked to endogenous proteins. % SecY, percentage of non-crosslinked SecY. **c**, NC100 with either a wild-type (WT) or defective (RR) signal sequence was expressed together with SecY_{68C} or the SecY(Δ P) mutant. Where indicated, cells were pre-treated with rifampicin before addition of BM. The samples were analysed by SDS-PAGE, followed by blotting with streptavidin–HRP conjugate

or with antibodies against Myc (detects NC100), SecY or TF. **d**, Nascent chains (NC) of different lengths, all with Cys 19, were expressed together with SecY_{68C} and crosslinked with CuPh₃. The samples were analysed by SDS-PAGE with or without prior reduction with β -mercaptoethanol (β -ME), followed by immunoblotting. Red arrowhead, crosslinked SecY and NC-tRNA. **e**, Nascent chains of different lengths with either wild-type (WT-NC) or defective (RR-NC) signal sequences were expressed together with the Δ P channel. BM was added to the cells and biotinylated proteins were detected by SDS-PAGE followed by blotting. The modification of p30 was quantified and normalized with respect to the value obtained without nascent chain expression (error bars, s.d.; $n = 3$). **f**, As in **e**, but spheroplasts were diluted into an iso-osmotic solution of xylitol and the initial, linear rate of turbidity decrease was determined (error bars, s.d.; $n = 3$). Spheroplasts were also analysed after treatment with rifampicin. The SecY(Δ P)+Rif sample showed an initial lag phase, which was ignored. **g**, As in **f**, but with dilution into iso-osmotic KCl containing valinomycin.

chains of 90 amino acids or longer were inserted into the SecY channel, as demonstrated by disulphide crosslinking (Fig. 2d). These chains almost completely blocked the permeation of BM through the SecY(Δ P) channel (Fig. 2e and Supplementary Fig. 7). Some reduction of BM permeability was also observed with chains of 60 to 80 residues, which were not inserted into the SecY channel. These chains are probably still targeted to the channel, as signal sequence mutants (RR) did not prevent BM permeation. Thus, the formation of a ribosome-channel junction may provide a barrier for BM, but complete blockage of permeation requires channel insertion of the nascent chain. Similar results were obtained when the permeation of xylitol was analysed by osmotic swelling and bursting of spheroplasts (Fig. 2f). The permeation of Cl^- ions through the SecY(Δ P) channel was inhibited only moderately by short nascent chains (Fig. 2g), but was reduced by about 90% by channel-inserted nascent chains, although not quite to the level seen with the wild-type channel. Chains with a defective signal sequence did not block Cl^- permeation significantly. Consistent with our observation that short chains do not block the permeation of the smallest molecules through the SecY(Δ P) channel efficiently, these cells grew significantly more slowly than those expressing channel-inserted chains (Supplementary Fig. 8). Wild-type SecY did not cause chain-length-dependent growth behaviour (Supplementary Fig. 8) and did not allow significant Cl^- uptake into spheroplasts regardless of nascent chain length, even when potentially counteracting pump activity was abolished by energy depletion (Supplementary Fig. 9). Collectively, these results show that the nascent chain itself provides an effective seal. Consistent with this notion, expression of the post-translational, SecA-dependent substrate proOmpA in the SecY(Δ P) mutant significantly reduced BM permeation, whereas a signal sequence mutant did not (Supplementary Fig. 10). The level of channel blockage was lower than that seen with a stalled ribosome-nascent-chain complex, probably because proOmpA occupies the SecY channel only transiently. It should be noted that some Cl^- permeation was observed *in vitro* with the wild-type SecY channel engaged in SecA-mediated translocation^{14,21} (see Supplementary Discussion).

Next, we tested whether the specific sequence of the nascent chain inside the channel affects the permeability for small molecules. We determined that residues ~39 to 60 of NC100 are inside the central pore (Fig. 2a), on the basis that approximately the last 40 residues are inside the ribosome and residues 19–34 are close to the plug domain on the periplasmic side of SecY (Supplementary Fig. 11). We then varied the sequence of the nascent chain inside the SecY channel, making it more hydrophilic or hydrophobic than the original NC100 sequence or replacing parts with stretches of glycines (Supplementary Fig. 12). All variants completely blocked BM and xylitol permeation through the SecY(Δ P) channel as effectively as the original NC100 chain (Supplementary Fig. 12 and data not shown) and the more hydrophobic chains were more potent in blocking Cl^- permeation (Supplementary Fig. 12). With the wild-type channel, little or no permeation was observed, regardless of which nascent chain was expressed. Thus, many different sequences of a translocating polypeptide can block the permeation of small molecules through the pore.

Next, we tested the role of the pore ring of SecY in sealing the resting channel. Replacement of just one of the six isoleucines in the pore by a glycine caused only a little BM permeation but the channel became progressively more leaky with increasing numbers of glycine substitutions (Fig. 3a). Alanine substitutions had less severe effects. BM permeation occurs through the resting channel because addition of rifampicin had no effect (Supplementary Fig. 13a). The pore-mutant channels were also permeable to xylitol (see below) and Cl^- ions (Supplementary Fig. 13b). In addition, the most severe pore mutants caused a strong growth defect (Supplementary Fig. 13c): the cells died immediately after induction of these SecY mutants (Supplementary Fig. 2). Cell death is probably caused by dissipation of the membrane potential because flow cytometry using the voltage-sensitive dye DiBAC4(3) showed that the membrane potential was decreased to about the same

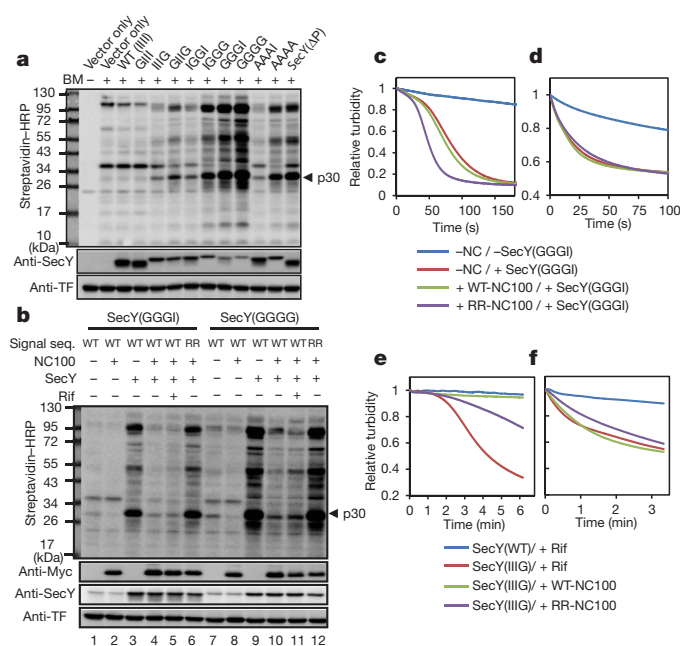


Figure 3 | Permeability in pore ring mutants. **a**, Wild-type SecY or pore mutants in which Ile 86, Ile 191, Ile 278 or Ile 408 (I in all positions) were replaced by Gly or Ala (G or A in the relevant positions) were expressed under the Tet promoter. Cells containing the SecY(Δ P) mutant or an empty vector were also analysed. After treatment of cells with BM, the samples were analysed by SDS-PAGE and blotting with streptavidin-HRP conjugate, or with antibodies against SecY or TF. **b**, NC100 with WT or RR signal sequence was expressed from the inducible arabinose promoter together with SecY pore mutants under the Tet promoter. Where indicated, rifampicin was added before BM. The samples were analysed by SDS-PAGE, followed by blotting with streptavidin-HRP conjugate or with antibodies against Myc, SecY or TF. Addition of rifampicin does not clear the mutant channels of nascent chains (lanes 5 and 11), probably because peptidyltransferase activity is compromised in these dying cells. **c**, **d**, As in **b**, but spheroplasts containing the GGGI pore mutant were diluted into iso-osmotic xylitol (**c**) or iso-osmotic KCl containing valinomycin (**d**), then the turbidity change was followed over time. **e**, **f**, As in **c**, **d**, but with the IIIIG pore mutant under the constitutive promoter with a GUG start codon. The cells were treated with rifampicin or induced for NC100 expression and then diluted into iso-osmotic xylitol (**e**) or KCl containing valinomycin (**f**).

extent as that seen at a 20 μM concentration of the ionophore carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP) (Supplementary Fig. 14). Taken together, these results show that the pore ring has an important role in maintaining the seal of the resting SecY channel and that a leaky channel is inconsistent with cell viability.

Finally, we investigated the role of the pore ring in sealing the active channel. Expression of NC100 blocked the permeation of BM through even the most severe SecY pore mutants (Fig. 3b, lanes 4 versus 3 and 10 versus 9) but expressing a signal sequence mutant (RR) of NC100 did not block permeation (Fig. 3b, lanes 6 and 12). By contrast, the permeation of the smaller molecules xylitol (Fig. 3c), Cl^- ions (Fig. 3d) or K^+ ions (Supplementary Fig. 15) was not prevented by expression of NC100 in a pore mutant containing three glycines. With a mutant containing only one glycine in the pore ring, NC100 expression blocked xylitol (Fig. 3e) but not Cl^- ions (Fig. 3f). The high permeability for Cl^- ions was maintained regardless of the sequence inside the SecY channel (Supplementary Fig. 16). Thus, a translocating channel requires isoleucines all around the pore ring to prevent small Cl^- ions from passing, whereas pore defects are tolerated for larger molecules (Supplementary Fig. 1).

Our results show that the wild-type SecY channel is sealed for even the smallest molecules, both in its resting state and when translocating a polypeptide. A simple model explains how this is achieved (Fig. 4). In

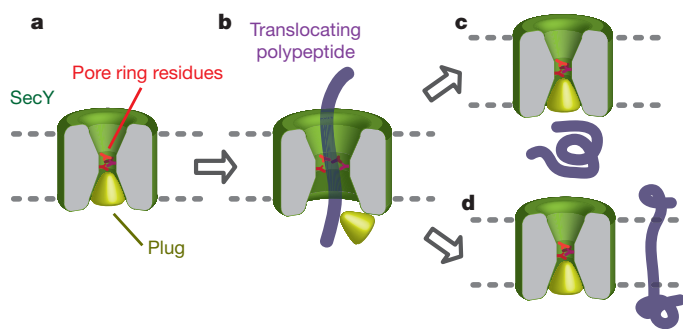


Figure 4 | Model for the maintenance of the membrane barrier by the SecY channel. **a**, SecY in the resting state: the plug is located in the centre, interacting with pore ring residues and sealing the channel. **b**, The plug is displaced during translocation; the pore ring forms a gasket-like seal around the translocating polypeptide chain to prevent the free flow of ions. **c, d**, When the polypeptide leaves towards the extracellular side (**c**) or sideways into lipid after the arrival of a hydrophobic transmembrane sequence (**d**), the plug returns to re-seal the channel.

the resting state, the plug is located in the centre of SecY, interacting with the pore ring residues and sealing the channel (Fig. 4a). During translocation, the plug is displaced and the pore ring forms a gasket-like seal around the translocating polypeptide chain to prevent the free flow of ions (Fig. 4b). The translocating chain itself serves as the major obstacle for small molecules; without it, the open pore allows many small molecules to pass. Whenever the polypeptide leaves the channel, either towards the extracellular side after termination of translocation (Fig. 4c), or sideways into lipid after the arrival of a hydrophobic transmembrane sequence (Fig. 4d), the plug returns and re-seals the channel. This mechanism applies to both co-translational and post-translational translocation. However, in co-translational translocation, the junction between the ribosome and the channel seems to provide an additional barrier for larger molecules, preventing metabolites and other cytosolic molecules from reaching the channel. Given the sequence conservation of the SecY and Sec61 channels, these principles may be universal. However, in prokaryotes, a tight seal is essential for cell viability, whereas in eukaryotes, the intracellular endoplasmic reticulum membrane may tolerate some leakiness^{22–24}, which could explain why Sec61 pore mutants in *Saccharomyces cerevisiae* have only minor growth defects²⁵.

METHODS SUMMARY

All strains and plasmids used in this study are listed and described in Supplementary Tables 1 and 2. The subunits of the heterotrimeric SecY complex were expressed from either the Tet promoter in the pTet vector (Figs 1 and 3a–d), or the endogenous *rplN* promoter on either the pACYC-SecYEG vector (Figs 2c, e–g and 3e, f) or the pRSY vector (Fig. 2d). The latter vector also contained the SRP components Ffh, 4.5S RNA and FtsY under their own promoters. NC100 and its variants were expressed from the arabinose promoter on the separate pBAD-NC100 plasmid, except in Fig. 2b, where both NC100 (from the arabinose promoter) and SecY (from the *rplN* promoter) were expressed from the same plasmid and the SRP components were expressed from another. Except in Fig. 2d, an *E. coli* strain was used in which chromosomal SecY is tagged at the carboxy terminus with a calmodulin-binding peptide (CBP). To measure permeability to BM, cell cultures were incubated with 0.4 mM BM, then quenched with 20 mM β -mercaptoethanol and lysed in SDS sample buffer. For cell fractionation, cells were converted into spheroplasts by treatment with EDTA and lysozyme in the presence of 18% sucrose. The spheroplasts were lysed by sonication and the membranes were separated from the cytosol by ultra-centrifugation. To measure permeability for other small molecules, spheroplasts were diluted 20-fold with iso-osmotic solutions and the absorbance was followed at 500 nm in a spectrophotometer. To disulphide-crosslink nascent chains with SecY, 0.25 mM copper(II)(1,10-phenanthroline)₃ (CuPh₃) was added to the culture directly, followed by quenching with 20 mM N-ethylmaleimide and lysis in SDS sample buffer.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 January; accepted 22 March 2011.

- Rapoport, T. A. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**, 663–669 (2007).
- van den Berg, B. et al. X-ray structure of a protein-conducting channel. *Nature* **427**, 36–44 (2004).
- Harris, C. R. & Silhavy, T. J. Mapping an interface of SecY (PrfA) and SecE (PrfG) by using synthetic phenotypes and *in vivo* cross-linking. *J. Bacteriol.* **181**, 3438–3444 (1999).
- Tam, P. C., Maillard, A. P., Chan, K. K. & Duong, F. Investigating the SecY plug movement at the SecYEG translocation channel. *EMBO J.* **24**, 3380–3388 (2005).
- Hamman, B. D., Hendershot, L. M. & Johnson, A. E. BiP maintains the permeability barrier of the ER membrane by sealing the luminal end of the translocon pore before and early in translocation. *Cell* **92**, 747–758 (1998).
- Liao, S., Lin, J., Do, H. & Johnson, A. E. Both luminal and cytosolic gating of the aqueous ER translocon pore are regulated from inside the ribosome during membrane protein integration. *Cell* **90**, 31–41 (1997).
- Simon, S. M. & Blobel, G. A protein-conducting channel in the endoplasmic reticulum. *Cell* **65**, 371–380 (1991).
- Saparov, S. M. et al. Determining the conductance of the SecY protein translocation channel for small molecules. *Mol. Cell* **26**, 501–509 (2007).
- Li, W. et al. The plug domain of the SecY protein stabilizes the closed state of the translocation channel and maintains a membrane seal. *Mol. Cell* **26**, 511–521 (2007).
- Zhang, W., Bogdanov, M., Pi, J., Pittard, A. J. & Dowhan, W. Reversible topological organization within a polytopic membrane protein is governed by a change in membrane phospholipid composition. *J. Biol. Chem.* **278**, 50128–50135 (2003).
- Bieker, K. L., Phillips, G. J. & Silhavy, T. J. The *sec* and *prl* genes of *Escherichia coli*. *J. Bioenerg. Biomembr.* **22**, 291–310 (1990).
- Derman, A. I., Puziss, J. W., Bassford, P. J. & Beckwith, J. A signal sequence is not required for protein export in *prlA* mutants of *Escherichia coli*. *EMBO J.* **12**, 879–888 (1993).
- Smith, M. A., Clemons, W. M. Jr, DeMars, C. J. & Flower, A. M. Modeling the effects of *prl* mutations on the *Escherichia coli* SecY complex. *J. Bacteriol.* **187**, 6454–6465 (2005).
- Dalal, K. & Duong, F. The SecY complex forms a channel capable of ionic discrimination. *EMBO Rep.* **10**, 762–768 (2009).
- Schierle, C. F. et al. The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the *Escherichia coli* periplasm via the signal recognition particle pathway. *J. Bacteriol.* **185**, 5706–5713 (2003).
- Nakatogawa, H. & Ito, K. The ribosomal exit tunnel functions as a discriminating gate. *Cell* **108**, 629–636 (2002).
- Nakatogawa, H. & Ito, K. Secretion monitor, SecM, undergoes self-translation arrest in the cytosol. *Mol. Cell* **7**, 185–192 (2001).
- Woolhead, C. A., Johnson, A. E. & Bernstein, H. D. Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Mol. Cell* **22**, 587–598 (2006).
- Ménétret, J. F. et al. Single copies of Sec61 and TRAP associate with a nontranslating mammalian ribosome. *Structure* **16**, 1126–1137 (2008).
- Becker, T. et al. Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* **326**, 1369–1373 (2009).
- Schiebel, E. & Wickner, W. Preprotein translocation creates a halide anion permeability in the *Escherichia coli* plasma membrane. *J. Biol. Chem.* **267**, 7505–7510 (1992).
- Le Gall, S., Neuhoof, A. & Rapoport, T. A. The endoplasmic reticulum membrane is permeable to small molecules. *Mol. Biol. Cell* **15**, 447–455 (2004).
- Heritage, D. & Wonderlin, W. F. Translocon pores in the endoplasmic reticulum are permeable to a neutral, polar molecule. *J. Biol. Chem.* **276**, 22655–22662 (2001).
- Roy, A. & Wonderlin, W. F. The permeability of the endoplasmic reticulum is dynamically coupled to protein synthesis. *J. Biol. Chem.* **278**, 4397–4403 (2003).
- Junne, T., Kocik, L. & Spiess, M. The hydrophobic core of the Sec61 translocon defines the hydrophobicity threshold for membrane integration. *Mol. Biol. Cell* **21**, 1662–1670 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Walter, H. Bernstein and G. Phillips for materials, D. Boyd for advice, C. Akey for discussions and C. Akey, A. Osborne and A. Salic for critical reading of the manuscript. The work was supported by a grant from the NIH (GM052586). T.A.R. is a Howard Hughes Medical Institute investigator.

Author Contributions E.P. performed the experiments and E.P. and T.A.R. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.A.R. (tom_rapoport@hms.harvard.edu).

METHODS

Bacterial strains and plasmids. All strains and plasmids used in this study are listed and described in Supplementary Tables 1 and 2. PCR reactions were performed with Phusion polymerase (New England Biolabs) or KOD polymerase (Novagen). Site-directed mutagenesis was performed using either the Quikchange (Stratagene) or Phusion (New England Biolabs) mutagenesis kits. All constructs were verified by sequencing. *E. coli* DH5 α strain was used for all cloning procedures.

For all experiments except cloning, we used *E. coli* strains lacking both Rmf and OmpT because Rmf is known to induce dimerization and inactivation of ribosomes at stationary phase²⁶ and the OmpT protease has been reported to proteolyse SecY²⁷. The Δ rmf Δ ompT strain was constructed from individual deletion strains by P1 transduction²⁸. The Δ rmf Δ ompT secY-CBP strain, in which the chromosomal secY gene is tagged at its C terminus with calmodulin-binding peptide (CBP), was constructed in the following way. We synthesized a 'CBP-RBS-Zeo' DNA cassette containing a CBP tag, a stop codon, a ribosome binding site and a zeocin resistance gene in that order. This cassette was amplified by PCR and electroporated into Δ rmf Δ ompT cells expressing Lambda red recombinase from the pKD46 plasmid²⁹. The resulting cells were selected on zeocin-containing agar plates (Invivogen). Incorporation of the cassette into the chromosome was verified by PCR, DNA sequencing and immunoblotting using CBP antibodies (Genscript). The Δ rmf Δ ompT secY-CBP strain was used throughout, except for experiments in Fig. 2d, where the Δ rmf Δ ompT strain was used. For all osmotic swelling and bursting experiments, we used a strain containing an additional deletion of the glpF gene because GlpF has been shown to allow permeation of polyalcohol sugars, including xylitol³⁰. The Δ rmf Δ ompT secY-CBP Δ glpF strain was generated by P1 transduction (see Supplementary Table 1) and did not show any xylitol permeability in osmotic swelling and bursting experiments.

The pTet-SecYEG plasmid, which expresses SecYEG under a tetracycline-inducible promoter, was made by PCR amplification of the SecE-SecY-SecG encoding sequence from the pBAD22-SecYEG plasmid⁹ and subsequent insertion of this sequence into the pTet vector. The pACYC-SecYEG plasmid, which constitutively expresses SecYEG, was constructed by inserting the same SecYEG coding sequence into pACYC184 after fusing it with a 200-base-pair DNA fragment containing the *E. coli* rplN promoter at its 5' end. The plasmid pBAD-NC100 for expression of a SecM-stalled nascent chain was generated as follows: we constructed the pBAD Myc-SecM vector by inserting a DNA fragment encoding a Myc epitope between the PstI and SacI sites of pBAD His/C (Invitrogen) and then inserting a fragment encoding the 17-residue SecM stalling sequence (FSTPVWISQAQGIRAGP) between the PstI and EcoRI sites. Then we synthesized a DNA segment by PCR encoding the DsbA signal peptide fused to a sequence from an unrelated protein and inserted it between the NcoI and PstI sites of the pBAD Myc-SecM vector. Other plasmids for different nascent chain lengths and sequences were made by modifying the pBAD-NC100 plasmid. More information is available in Supplementary Table 2.

Cell culture and protein expression. Unless otherwise indicated, *E. coli* cells were grown and induced as follows. Cells harbouring the indicated plasmids were picked from freshly transformed colonies and inoculated into LB medium supplemented with appropriate antibiotics (100 μ g ml⁻¹ for ampicillin and 50 μ g ml⁻¹ for chloramphenicol). Cultures were grown at 37 °C to log phase (optical density at 600 nm (OD₆₀₀) of about 0.4–0.6) before induction. To induce SecYEG from the pTet vector (Figs 1 and 3a), 200 ng ml⁻¹ anhydrotetracycline (aTet) was added for 30 min. To overexpress nascent chains, 0.2% arabinose was added for 1 h except in Figs 2b and 2d, where the induction was for 2 h. When both SecYEG (from pTet) and a nascent chain were co-expressed (Fig. 3b–d), 0.07% arabinose was added to *E. coli* cultures at log phase, then 200 ng ml⁻¹ aTet was added after 15 min and induction was continued for an additional 40 min. Where indicated, 100 μ g ml⁻¹ rifampicin was added to cultures for 1 h at 37 °C.

SecY was expressed from the endogenous rplN promoter on either the pACYC-SecYEG vector (Figs 2c, e–g and 3e, f) or the pRSY vector (Fig. 2d). The latter also encodes rare-codon tRNAs and the SRP pathway components (Ffh, 4.5S RNA and FtsY) from their endogenous promoters. For the experiment in Fig. 2b, a fused vector (pBAD-NC100/SecYEG) was used, encoding both NC100 and SecY, together with pr2HQ4, a vector overexpressing the SRP components and rare-codon tRNAs. Details about these plasmids and their construction are given in Supplementary Table 2. When a nascent chain was intended to saturate the SecY channels, vectors contained a GUG start codon for SecY instead of AUG (Figs 2b, c, e–g and 3e, f). Saturation of GUG-SecY with the NC100 chains was not affected by plasmid combinations or by overexpression of the rare-codon tRNAs and the SRP components.

In vivo biotinylation and subcellular fractionation. *E. coli* cells were grown to log phase at 37 °C in LB medium and induced as described above. After an aliquot of the culture was taken, 0.4 mM biotin-PEG₂-maleimide (BM; Pierce) was added to the culture directly for 30 min at room temperature (23 °C). When a nascent

chain was overexpressed, the incubation was shortened to 15 min. After quenching the reaction with 20 mM β -mercaptoethanol and a further incubation of 15 min on ice, the cells were collected by brief centrifugation. For SDS-PAGE analysis, the same numbers of cells were lysed in SDS sample buffer and loaded on a gel (0.1 OD₆₀₀ per lane). For subcellular fractionation, the cells were re-suspended in ice-cold spheroplasting buffer (100 mM Tris-HCl, pH 8.0, 18% sucrose). Conversion of cells into spheroplasts was carried out by addition of 2 mM EDTA and 0.1 mg ml⁻¹ hen egg lysozyme and incubation for 10 min on ice. The spheroplasts were sedimented by centrifugation (9,000 r.p.m., 10 min) and the supernatant containing periplasmic proteins was removed. The spheroplasts were re-suspended in 100 mM Tris-HCl, pH 8.0, 150 mM NaCl and lysed by sonication. The lysate was subjected to centrifugation for 1 h at 51,000 r.p.m. in a TLA100.3 rotor (Beckman) to pellet the membranes. The supernatant (cytosolic proteins) and the membranes were analysed by SDS-PAGE.

Osmotic swelling and bursting experiments. Cells were placed on ice for a few minutes and harvested by centrifugation at 5,000 r.p.m. for 7 min. After re-suspension in one-fifteenth culture volume of 20 mM Tris-HCl, pH 7.2 and 18% sucrose (0.619 mol kg⁻¹), the cells were converted to spheroplasts by addition of 2 mM EDTA and 0.1 mg ml⁻¹ lysozyme at 4 °C. When nascent chains were expressed, 4 mM MgSO₄ was added 3 min after initiation of spheroplasting to avoid potential adverse effects of EDTA on ribosome–nascent chain complexes. Efficient (>95%) spheroplasting was verified by phase contrast microscopy and by immediate lysis on dilution into water (confirmed by measuring turbidity). To determine permeability for various small molecules, the spheroplast suspension was mixed rapidly at room temperature with a 19-fold volume of an iso-osmotic solution of xylitol (0.616 mol kg⁻¹) or KCl (0.342 mol kg⁻¹). Absorbance at 500 nm was recorded in a spectrophotometer. The time between mixing and measurement was about 5 s. To obtain more accurate iso-osmotic solutions, osmotic coefficients were taken into account (1.05 for sucrose, 0.90 for KCl and 1.00 for xylitol). For measurement of chloride permeability, 10 μ M valinomycin was included in the KCl solution. When the rate of turbidity decrease was determined, initial slopes were calculated by linear regression. These numbers were normalized with respect to the initial turbidity.

In vivo expression and crosslinking of nascent chains. To disulphide-crosslink a nascent chain with SecY, 0.25 mM CuPh₃ was added directly to the induced culture. After gentle rocking at room temperature for 10–20 min, the cells were collected by centrifugation and re-suspended in ice-cold TMP100 buffer (50 mM Tris-acetate, pH 7.2, 25 mM Mg(OAc)₂, 0.1 M KOAc) containing 20 mM N-ethylmaleimide. The cells were lysed by sonication and the samples analysed by non-reducing SDS-PAGE and immunoblotting. Where indicated, the samples were treated with 2% β -mercaptoethanol or 0.2 mg ml⁻¹ RNase A before loading them onto the gel.

Testing for SRP-dependence of nascent chain insertion. WAM121 *E. coli* cells (Supplementary Table 1) were transformed with pRARE/SecYEG and pTac-NC100 (Supplementary Table 2). Cells from a saturated culture, grown in LB medium supplemented with 0.1% arabinose, were washed three times with RM medium (1 \times M9 salt, 1 mM MgSO₄, 2% casamino acids and 0.2% glucose) supplemented with ampicillin and chloramphenicol and then inoculated into fresh RM medium for 3.5 h at 37 °C. Incubation was then continued in either the absence or presence of 0.1% arabinose. After an additional 1 h, overexpression of the nascent chain was induced for 2 h by addition of 1 mM isopropyl thiogalactopyranoside (IPTG). *In vivo* crosslinking experiments were performed as above.

Mal-PEG modification of nascent chains. pBAD-NC100 plasmids encoding single cysteines at various positions were transformed into Δ rmf Δ ompT cells harbouring the pRARE plasmid. After induction of nascent chain expression, the cells were harvested and lysed by sonication in TMP100 buffer. Mal-PEG ((methyl-PEG)₁₂-PEG₄-maleimide, Pierce) was prepared as a 125 mM solution in dimethyl sulphoxide (DMSO) and added to the cell homogenate at 2 mM final concentration. After incubation for 1 h at room temperature, the reaction was stopped by addition of 33 mM N-ethylmaleimide. The samples were subjected to SDS-PAGE and analysed by immunoblotting with antibodies against Myc.

Flow cytometry. To monitor the membrane potential of cells, SecY was induced for the indicated time period, then cells were diluted to $\sim 5 \times 10^6$ cells ml⁻¹ in phosphate buffered saline containing 2 μ M bis-(1,3-dibutylbarbituric acid) trimethine oxonol (DIBAC4(3), Invitrogen). After a few minutes, the cell suspension was injected into a flow cytometer (FACS calibre, BD) for analysis. The flow rate was set at 'low'. Signals from cells were readily distinguishable from background signals as a separate population in forward vs. side scattering (FSC versus SSC) plots. This population was gated to exclude background events from further analysis. The DIBAC4(3) signal was measured in the green fluorescence channel (FL1-H). A total of 50,000 events were counted in each experiment.

SDS-PAGE, immunoblotting and densitometry analysis. SDS-PAGE was performed using Bis-Tris gels (Invitrogen) with MES-SDS running buffer. Images of immunoblots were taken with a CCD-based device (Fujifilm LAS-3000) and a standard ECL reagent. Image J software was used for densitometry analysis.

26. Wada, A., Yamazaki, Y., Fujita, N. & Ishihama, A. Structure and probable genetic location of a "ribosome modulation factor" associated with 100S ribosomes in stationary-phase *Escherichia coli* cells. *Proc. Natl Acad. Sci. USA* **87**, 2657–2661 (1990).
27. Akiyama, Y. & Ito, K. SecY protein, a membrane-embedded secretion factor of *E. coli*, is cleaved by the OmpT protease in vitro. *Biochem. Biophys. Res. Commun.* **167**, 711–715 (1990).
28. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
29. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
30. Heller, K. B., Lin, E. C. & Wilson, T. H. Substrate specificity and transport properties of the glycerol facilitator of *Escherichia coli*. *J. Bacteriol.* **144**, 274–278 (1980).

CAREERS

TURNING POINT Hydrodynamics award leads biophysicist to wind energy **p.245**

POSTDOCS UK survey says universities offer too little career support **p.245**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

STOCKTREK IMAGES



Geoscientists are needed in areas such as Ethiopia's Danakil Depression to study tectonic plates.

GEOSCIENCES

Earth works

There's good news for aspiring geoscientists. Job opportunities at all career stages are on the rise.

BY SID PERKINS

Mention geoscience and people often imagine trekking to far-flung regions to hammer rocks. But the discipline offers a wide range of opportunities beyond this. "There's room for those who love field work, and there's room for those who don't," says Eric Calais, a geophysicist at Purdue University in West Lafayette, Indiana, who has spent time in the field studying the movements and deformations of Earth's tectonic plates in Ethiopia, Siberia and Indonesia. Calais recently left the lab again — this time as science adviser to the United Nations Development Program's mission to quake-torn Haiti, where he is helping to develop public-safety policy and working with local scientists, government officials and international aid workers to build a national agency for seismic risk reduction. "Data analysts, computer modellers — geoscience needs all types of researchers," he notes.

Some geoscientists — a term encompassing geologists and geophysicists — study the composition, structure and other physical aspects of Earth, and its geological past and present, to search for and extract natural resources such as oil, gas and minerals. Others help to preserve and clean up the environment. Yet others are expert in fields such as atmospheric chemistry, oceanography and deep-earth mineralogy. Geoscientists find jobs in environmental services, scientific and technical consulting, government and academia — and even in high finance, where insurance companies rely on them to help assess long-term risk due to climate change, earthquakes, hurricanes and other natural disasters.

And while a few jobs in geoscience — particularly those in the private sector — require only a bachelor's degree, for most positions, education to master's-degree level or higher will be needed. According to the American Geological Institute (AGI), headquartered in

Alexandria, Virginia, more than two-thirds of geoscience PhDs find their first jobs in academia, with the remainder landing positions in government and industry (see 'Where geoscientists work').

Despite the recent economic downturn, job prospects for geoscientists are excellent and are set to get even better. The latest data from the US Bureau of Labor Statistics (BLS; www.bls.gov) suggest that in 2018 there will be some 323,000 positions for geoscientists in the United States, about 23% more than in 2008. "If you have a degree, you'd have to be dead or dead lazy to not get a job," says Christopher Keane, AGI's director of technology and communications.

Yet there's even more room for optimism, says Leila Gonzales, a workforce analyst with the AGI. The BLS data only project changes in the number of jobs, she notes. They don't account for increased opportunities in the United States due to retirements, which will strike government, academia and the private sector alike during the next 10–15 years. This window of opportunity is the result of a collapse in commodity prices in the mid-1980s that was followed by a precipitous decline in geoscientists' salaries, which in turn triggered a drop in academic enrolments. In the next decade or so, many of those graduates from the early 1980s will be nearing the ends of their careers, boosting job availability upon their retirement.

And the number of retirements could be substantial, says Gonzales. About 12% of today's geoscientists are expected to retire by 2018, meaning that net job availability for geoscientists in the United States should have increased between 2008 and 2018 by around 35%, she says.

GOING GLOBAL

Demand for geoscientists is expected to rise worldwide, not just in the United States. Statistics compiled by the AGI for the International Union of Geological Sciences indicate that the three regions that produce the most geoscience graduates each year — the United States; Europe and Russia combined; and China — aren't even meeting their own domestic needs, much less global requirements.

The number of geoscience degrees granted in the United States each year (about 1,060 master's and 650 PhD degrees) has remained relatively stable for the past 15 years or more, says Gonzales. That steady, yet inadequate, pace of new entries into the field, paired with the wave of retirements during the next 10–15 years, almost ►

► guarantees that demand for skilled geoscientists can only increase.

In the past, about half of the global demand for geoscientists has come from the United States. But now and in the near future, rapid economic expansion in India, China and the rest of the developing world is expected to boost international demand for geoscience graduates. Highly skilled geoscientists will be needed to help identify and develop oil, gas and mineral resources, as well as to help recognize and ameliorate natural and man-made environmental hazards in these developing markets.

Until now, the United States and Europe have been able to import geoscientists from other countries to meet demand, while emerging economies have typically struggled to retain their own talent. Areas with a desperate need for geoscientists — South Africa, Nigeria, and central and eastern Europe, among others — continue to lose new graduates in search of higher wages and additional education in North America and western Europe.

By contrast, the export of geoscientists from the developing world provides incredible opportunities for fledgling geoscientists from developed countries to gain experience in those regions — especially if they are willing to work under difficult conditions in remote areas that may be politically unstable.

If recent trends continue, many of the new positions will be outside the United States. Between 2007 and 2009, the proportion of geoscience PhDs working outside the United States rose from 10% to 12%. The key to career success may indeed be mobility, Keane, Gonzales and their colleagues contend. “If you’re willing to move around, there’s great opportunity,” says Keane.

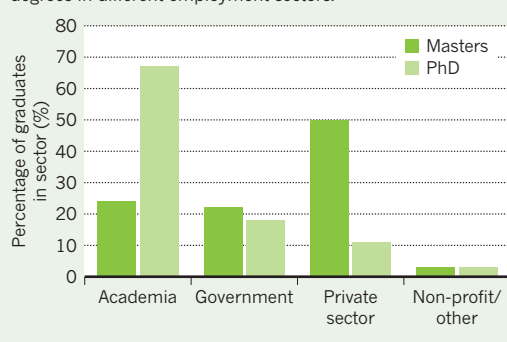
IN HIGH DEMAND

Despite the growing desire of many countries to shift to renewable sources of energy, economic growth in the foreseeable future will continue to bolster demand for fossil fuels and for the geoscientists needed to find them. Even in the oil and gas companies, which typically offer some of the highest salaries, the projected supply of new talent won’t come close to meeting demand. By 2030, the petroleum industry is likely to have at least 13,000 unfilled jobs, according to data compiled by the AGI.

One of the geoscience employment sectors poised for growth is consultancy, according to the BLS, which includes it in the “professional,

WHERE GEOSCIENTISTS WORK

Percentage of recently graduated geoscientists with advanced degrees in different employment sectors.



scientific and technical services” sector. Geoscience research consultants in the United States can work either independently or as employees of private contractors. The BLS predicts that between 2008 and 2018 the sector will grow by more than 50%, even without counting retirements. This growth will provide opportunities for geoscientists equipped with a strong set of fundamental skills, particularly postdocs with several years’ experience who elect to leave academia, and mid-career researchers who choose to leave government positions. Projected growth in this sector is largely driven both by increased demand in the private sector and by the US government’s increasing inclination to outsource work rather than hire new employees.

There will be some jobs for geoscientists in government, however. The number of geoscience posts across federal, state and local government in the United States is expected to have risen by about 7.7% by 2018, says the BLS. Growth will be less at the federal and local levels (around 6.9% and 7.1%, respectively) but stronger at the state level (about 8.7%), with many of the positions arising in agencies tasked with implementing regulations on environmental issues such as water quality.

Many of today’s senior geoscientists were trained as specialists in relatively narrow disciplines, but in future, most demand will be for researchers who have been trained to appreciate the interdisciplinary nature of the Earth sciences.

For example, a geoscientist studying toxic algal blooms in coastal waters will need to understand how rivers, streams and groundwater become tainted by fertilizer, how quickly those pollutants reach the sea, how quickly the algae grow under different environmental conditions and how those growth rates might vary under climate change — an understanding that would come from training in geology, oceanography and biogeochemistry.

Geoscientists who find employment in the private sector are likely to work in teams of people with diverse backgrounds, which means that a broad education and experience — as well as good interpersonal skills and the ability to

communicate effectively — must-haves in the workplace.

Devising solutions to many of today’s most pressing problems will require knowledge spanning several fields, says Jeff Gaffney, an atmospheric chemist at the University of Arkansas in Little Rock. To evaluate the potential costs and benefits of switching from fossil fuels to biofuels, for example, researchers must consider issues such as air quality, water quality, agricultural development and the effects of fertilizers and pesticides needed to grow the crops that produce the biomass feedstock. At a minimum, researchers must be familiar with the jargon used by team members of other disciplines. “Otherwise, everyone’s talking a different language,” Gaffney notes.

But in addition to becoming more interdisciplinary, geoscience is becoming more technical, especially in certain subdisciplines such as atmospheric chemistry and climate change, where computer modelling has become a mainstay. A strong background in mathematics, particularly in statistics and in computer simulation techniques, will enable researchers to understand and interpret models used to simulate the behaviour of everything from ocean currents to the material properties of minerals deep in Earth’s mantle.

COMPLEX SYSTEMS

“Many earth-science systems are very complicated,” Gaffney says. “Researchers need to understand not only the mathematics of the models they’re using but the basic fundamentals of the physical and chemical processes that they’re trying to simulate,” he says. And that’s especially true for models used in analyses that could have substantial effects on public policy, such as those related to climate change or to the potential costs and benefits of alternative fuels.

Computer skills will be indispensable for almost all positions in geoscience, and students who have acquired experience in data analysis, digital mapping, remote sensing and the ubiquitous Global Positioning System and geographical information systems will be best prepared to enter the job market. Computer programming skills will also be essential, not only for developing or enhancing detailed computer models but for designing sensors, equipment and data-collection networks. Geoscientists will either need to develop these skills or know enough about them to work with specialist colleagues carrying out those tasks.

For those willing to get interdisciplinary training, the future looks bright. Someone entering an undergraduate geoscience programme today and then pursuing a master’s or PhD degree will enter a job market flush with opportunities — especially if they are willing to relocate. ■

Sid Perkins is a freelance science writer based in Crossville, Tennessee.



“There’s room for those who love field work and there’s room for those who don’t.”

Eric Calais

TURNING POINT

John Dabiri

A. SUMMA/MACARTHUR FOUNDATION

John Dabiri, a biophysicist at the California Institute of Technology (Caltech) in Pasadena, was named a MacArthur Fellow last September for his work on the hydrodynamics of jellyfish swimming. The award brings a US\$500,000 research grant over five years.

What does this fellowship mean to you?

I don't get a lot of external feedback, because my field is small, and the award signals that at least some people out there appreciate my work. I hope it will encourage others to do unconventional research.

What are some applications of your findings?

The flow dynamics of jellyfish propulsion are similar to cardiac blood flow. In the heart, oxygenated blood enters the left ventricle and forms a vortex; that structure is important for efficient heart function. My colleague Mory Gharib, a bioengineer at Caltech, suggested looking at cardiac blood flow as an index for heart failure. Essentially, you would measure that vortex and determine whether the heart was approaching dysfunction — ideally, earlier than is possible now. We can use jellyfish hydrodynamics to model the heart.

What are you working on at the moment?

I'm making a foray into wind energy. Most wind turbines rotate on a horizontal axis, but I'm researching vertical-axis turbines. Conventional turbines can accept wind from only one direction, but these can take it from anywhere. The only problem is that they convert less wind into energy. But the total energy output of a group of vertical-axis turbines can be superior to the status quo.

Is this linked with your jellyfish work?

It's related more generally to fish swimming. We had done some modelling to work out how to arrange vertical-axis turbines for greatest efficiency, and it occurred to me that we could gain information from how fish swim. When fish arrange themselves in schools, they can interact with vortices created by the fish next to them. The question is, do certain arrangements minimize energy expenditure, and are ten fish more efficient on average than one? As my team and I set up our wind farm, we applied our knowledge of flow dynamics and fish schools to maximize the energy that turbines take out of the air.

How is this project progressing?

Last September we received a grant from



the Gordon and Betty Moore Foundation in Palo Alto, California, to buy a parcel of land and create a research wind farm with 42 9-metre-tall wind turbines. This is an excellent opportunity to test our modelling and prove our ideas in the real world — we don't have to rely on a wind tunnel or computer simulations.

What advice do you give to new junior faculty members?

Academia can be a roller coaster of good and bad news. The same year that the US National Science Foundation rejected my grant application for the third and final time, the work that they had turned down was published on the cover of *Nature*. And a week after winning the MacArthur fellowship, I had a paper related to the relevant work rejected by another journal. Don't let praise or criticism get to you — it's a weakness to get caught up in either one.

What achievement are you most proud of?

In my field there are few black researchers, so having the opportunity to work and teach at Caltech and to demonstrate the promise of under-represented groups is exciting for me. I hope that I can inspire young students in the African American community and help them to realize that they have options.

Do you have a secret for scientific success?

I work hard and pray a lot, and I keep in mind that at the end of the day, there's a lot that's out of my control. You can't control the reviewer who isn't going to like your study, or whether your great idea will get scooped by another lab. But you can control whether you show up and put in the effort. ■

INTERVIEW BY KAREN KAPLAN

GRADUATE STUDENTS

US applications rise

Overseas applications to US graduate institutions increased by 9% in 2010–11, says a report released on 9 April by the US Council of Graduate Schools (CGS) in Washington DC. The 2011 CGS *International Graduate Admissions Survey* indicated that applications from China were up by 18% and those from the Middle East and Turkey rose by 12%, marking the sixth year in a row with double-digit increases from all three regions. Applications from India and South Korea rose by 7% and 2% respectively, following large declines in 2008–09. All fields of study showed growth in 2010–11, led by 12% increases in overseas applications to physical and Earth sciences and engineering. Life-sciences applications rose by 8%, the field's largest gain since 2007.

UNITED KINGDOM

Postdocs need support

UK postdoctoral researchers need better career advice, finds a survey of 776 postdocs by the Royal Society of Chemistry and the Institute of Physics, both in London. *Mapping the Future: Physics and Chemistry Postdoctoral Researchers' Experiences and Career Intentions*, released on 14 April, says that less than half have regular appraisals, and impartial advice about non-academic careers is scarce. Female chemists were most likely to have doubts about staying in academia, and male physicists spent the most time as postdocs: an average of 3–4 years, compared with 1–2 years for women. The problems should be tackled by universities, funding bodies and the Concordat Strategy Group, a UK postdoc-support organization, says the report.

CANCER THERAPEUTICS

Massachusetts openings

Blueprint Medicines, a start-up cancer-therapeutics company based in Cambridge, Massachusetts, is to recruit up to 20 researchers. In April, the firm secured US\$40 million in venture capital, the largest first-round financing deal in 2011 so far for a life-sciences company, says the US National Venture Capital Association in Arlington, Virginia. Chris Varma, co-founder of Blueprint, says that the start-up is seeking early-career and established medicinal chemists, cell and structural biologists, and bioinformaticians to help discover genomics-based therapies.

THE UNIVERSE REEF

Up, up and away.

BY TOBIAS BUCKELL

Jackson buckles his leathers tighter and pulls on a fur. “The height causes that cold,” he shouts. “We’re like mountaineers!”

I want to flee the bitter cold and escape the wind, which seems to pierce my skin and scrub my bones. But I don’t want to miss seeing the Stone Table with my own eyes.

We’re standing on a catwalk that juts out from the skin of the airship and connects to the giant propellers on either side of the mid-belly area. The large blades are still, as the captain has found us a current of air. To save fuel we’re drifting, occasionally correcting our course when engines to the rear of the whale-like lighter-than-air machine roar to life.

Underneath our feet: a mile of air. And then below that is the brown, rippling mass of the Reef.

Once upon a time, there was no Reef. The world looked vastly different. There are preserved pictures of this time, spirited away from the museums before they fell to the reef. But more than we can ever remember will always be trapped where they were stored in great cities of legend like Paris, London or Washington, where great men once had grand adventures.

What history, legend and archaeologists agree on was that something split the sky asunder. And the debris that rained down from above was not just meteorite. Something else struck the earth and the water.

It was a reef. Tiny beings deposited tiny skeletons that were built on and ossified until an entire ecosystem accreted. And more alien organisms, hidden away inside the remains of the rocks that fell, flowered around the reef. The alien flora marched across the ground, but left the oceans alone.

The Reef ate cities as it spread across the world, seeking out metal with a hunger than no one could quench.

Our ancestors fought it. Men from a different time, from those old nations, with those old technologies, unleashed hell upon the Reef. And sometimes they would slow it. Sometimes they would even kill it.

But it always came back.

It was the Reef. Inexorable and implacable. It reshaped the world.

Jackson Smithik is an adventurer. Those thick dreadlocks of his are growing grey with age, and his face is leathery from exposure to the Sun. He was the first person, post-Collapse,



to sail across the Atlantic, back to skirt the Reef-choked coasts of Africa, down to the cape, and then sail out to make contact with the Indian and then Pacific Islands.

Because it was only the smaller islands that survived the Reef, isolated by the ocean and far from the Reef’s continental creep.

Seventy years after his teenage captaincy and exploits, Smithik’s Jamaican Clippers roam the world’s oceans, connecting the world. And now, thanks to the advances of steam and steam-powered airships by the Icelandic Empire, Smithik Transport ships explore the skies.

“There it is,” Smithik shouts. A grey wall rises out of the Reef, which covers what was once the land of South America. And above the Stone Table rises The Tower.

I follow the bulk of the structure. It is too much. It is a mountain in the distance that tapers off into a needle that pierces the clouds. And keeps going.

“This is what the Reef was for,” Smithik yells into my ear. His eyes gleam.

Over some strong Blue Mountain coffee, back inside where it was warm, Smithik tells me: “Pre-Reef scientists had a theory called panspermia: they believed life on Earth was caused by small

organisms aboard comets thrown from collisions in other solar systems crashed down to seed life here, and maybe elsewhere.

“So a follow up infection, that’s not so hard to believe, yeah?”

I nodded and kept notes. I’d been paid to document his first trip to actually step onto Stone Table since Smithik’s adventurers had found it and reported back.

We land on the massive Reef-grown artificial stone structure and moor the airship. The joint Japanese and Hawaiian expedition group, and the Icelandic scientists who’d beaten them there, greet us.

Pictures are taken with the excited scientists and the man who had funded the first expedition to Stone Table, found when Caribbean telescopes had spotted the slowly self-assembling tower to space.

“We can’t say if the Reef is designed to create The Tower, programmed by some distant intelligence,” the scientists tell Smithik as I scribble. “It could just be the way the Reef reproduces, creating a way to fling its spores back into space.”

“But the Stone Table, and the grooves in The Tower, they’ll allow us to climb it with a machine into space? Doesn’t that prove it’s made for intelligence?” Smithik asks.

“Sometimes nature builds something something else can use. Maybe it’s hoping we’ll spread Reef spore as we use this to get into space.”

“As if we were bees,” Smithik nods.

Late in the night I stand with Smithik at the base of The Tower, looking up at the night sky. “Pre-Reef men once walked on the Moon,” he says.

“And you think we’ll go back?”

“Whether we’re part of some galactic ecosystem that the Reef is just a spore of, or whether something designed it, the more we explore out there the more we’ll understand what happened down here.”

The great adventurer died that night. But his spirit lives on in the Smithik Ascender, a plan by the international scientists to build a steam-powered climber that will ascend The Tower to space.

What we will find, no one knows. ■

Tobias Buckell was born in the Caribbean but now lives in Ohio. He is the author of multiple novels and can be found online at www.TobiasBuckell.com.

➔ **NATURE.COM**
Follow Futures on
Facebook at:
go.nature.com/mtoodm

entists had a theory called panspermia: they believed life on Earth was caused by small

Structure and function of a membrane component SecDF that enhances protein export

Tomoya Tsukazaki^{1*}, Hiroyuki Mori^{2*}, Yuka Echizen¹, Ryuichiro Ishitani¹, Shuya Fukai^{3,4}, Takeshi Tanaka⁵, Anna Perederina⁶, Dmitry G. Vassilyev⁶, Toshiyuki Kohno⁵, Andrés D. Maturana⁷, Koreaki Ito⁸ & Osamu Nureki¹

Protein translocation across the bacterial membrane, mediated by the secretory translocon SecYEG and the SecA ATPase^{1–4}, is enhanced by proton motive force^{5,6} and membrane-integrated SecDF^{7–9}, which associates with SecYEG. The role of SecDF has remained unclear, although it is proposed to function in later stages of translocation as well as in membrane protein biogenesis^{4,10–13}. Here, we determined the crystal structure of *Thermus thermophilus* SecDF at 3.3 Å resolution, revealing a pseudo-symmetrical, 12-helix transmembrane domain belonging to the RND superfamily and two major periplasmic domains, P1 and P4. Higher-resolution analysis of the periplasmic domains suggested that P1, which binds an unfolded protein, undergoes functionally important conformational changes. *In vitro* analyses identified an ATP-independent step of protein translocation that requires both SecDF and proton motive force. Electrophysiological analyses revealed that SecDF conducts protons in a manner dependent on pH and the presence of an unfolded protein, with conserved Asp and Arg residues at the transmembrane interface between SecD and SecF playing essential roles in the movements of protons and preproteins. Therefore, we propose that SecDF functions as a membrane-integrated chaperone, powered by proton motive force, to achieve ATP-independent protein translocation.

The functional importance of SecDF in protein translocation was previously shown *in vivo*^{7,9,14}: the SecD- and SecF-deficient *Escherichia coli* mutant *secD1* (producing negligible levels of SecD and SecF and exhibiting cold-sensitivity (Cs) for growth) is severely defective in protein export at any temperature. In *E. coli*, SecD and SecF (EcSecD and EcSecF respectively) are encoded by the *secD* operon (*yajC-secD-secF*) and form a tight complex⁸. Each protein contains six transmembrane (TM) segments, belonging to the resistance modulation and cell division (RND) superfamily, and also a large periplasmic domain^{15,16}.

SecDF from *Thermus thermophilus* HB8¹⁷ is composed of a single polypeptide chain (TtSecDF, molecular mass 80.5 kDa) containing 12 predicted TM regions and six periplasmic regions (P1–P6), among which P1 (TtSecDF_{31–263}) and P4 (TtSecDF_{469–559}) are large enough to form distinct domains (Supplementary Fig. 1). We solved the crystal structure of full-length TtSecDF at 3.3 Å resolution (Fig. 1a–c, f, Supplementary Fig. 2 and Supplementary Table 1), as well as the crystal structure of P1 (TtSecDF_{36–263}) at 2.6 Å resolution and the nuclear magnetic resonance (NMR) solution structure of P4 (TtSecDF_{481–557}) (Fig. 1d, e and Supplementary Table 2). The TtSecDF structure contains twelve TM α -helices with both termini facing the cytoplasm and P1 and P4 protruding into the periplasm (Fig. 1a, b). The N-terminal (TM1–6) and C-terminal (TM7–12) halves are assembled pseudo-symmetrically, with TM4 and TM10 forming the primary interface between SecD and SecF (Fig. 1c). The TM helices are membrane-embedded, except for the ~10 Å extensions of TM2 and TM8 that tether P1 and P4

respectively (Fig. 1a). The tilted TMs 4, 5, 6, 10, 11 and 12 are curved near the cytoplasmic surface.

The NMR structure of P4 revealed a ferredoxin-like fold, consisting of four β -strands flanked by two α -helices (Fig. 1e). P1, protruding by ~40 Å from the membrane, is divided into head (~25 Å periplasmic protrusion) and base subdomains (Fig. 1a, d). The base subdomain of P1 (residues 36–111 and 251–263) is structurally homologous to P4; both form pseudo-symmetrical, eight-stranded, anti-parallel β -sheets that cover the TM region (Fig. 1a, b). The crystal structure of isolated P1 (called the I form) revealed a ~120° rigid-body rotation of its head subdomain towards the periplasmic space, as compared to the orientation observed for P1 in full-length SecDF (called the F form) (Supplementary Fig. 3). Thus, SecDF can assume at least two different conformations, F and I, with distinct P1 configurations (Fig. 1f, g). For this conformational change, the two loops connecting the head and the base could act as a hinge (Fig. 1a, d), allowing the head to swing around the base. We built a model of the ‘full-length I form’ by superimposing the base subdomain of the isolated P1 structure onto that of the full-length SecDF. The *in vivo* occurrence of the two conformations was supported by disulphide crosslinking experiments, which further suggested that the conformational transition of SecDF is functionally important (Supplementary Fig. 4a–e and Supplementary Discussion).

Despite the vital importance of SecDF *in vivo*, its specific role in translocation has not been defined. We examined whether SecDF functions in the steps subsequent to the SecA- and ATP-dependent initiation of translocation. A pro-outer membrane protein A (pro-OmpA) translocation intermediate was generated by subjecting ³⁵S-labelled pro-OmpA(L59)¹⁸ to translocation reactions in the presence of SecA and ATP, with either wild-type inverted membrane vesicles (IMVs) or SecD- and SecF-deficient IMVs from the *secD1* (Cs) mutant. This intermediate possessed a 59-residue disulphide-linked loop acting as an obstacle to translocation through the Sec translocon (Fig. 2b). ATP was then depleted from the intermediate-bearing IMVs by glucose-hexokinase reaction cycles (Fig. 2a, lanes 1 and 6). The addition of dithiothreitol to the wild-type IMVs restored translocation that could be supported by a proton motive force (PMF) even in the absence of ATP (compare Fig. 2a, lanes 2 and 4). In contrast, the *secD1* IMVs did not support any further translocation of the intermediate, regardless of the presence of a PMF (Fig. 2a, lane 7). Instead, these IMVs tended to lose the intermediate (Fig. 2a, lanes 7 and 8), probably owing to backsliding of the polypeptide¹⁹. In the presence of excess ATP, the completion of pro-OmpA(L59) translocation was not affected by either the *secD1* mutation or the lack of PMF (Fig. 2a, lanes 13, 15, 18 and 20). These results show that SecDF is required for the later, ATP-independent steps of translocation driven by the PMF, whereas the PMF is thought to accelerate even the initial stage of translocation (Supplementary Discussion). In addition, IMVs bearing inactive

¹Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0032, Japan. ²Institute for Virus Research, Kyoto University, Sakyo-ku, Kyoto 606-8507, Japan. ³Structural Biology Laboratory, Life Science Division, Synchrotron Radiation Research Organization, The University of Tokyo, Tokyo 113-0032, Japan. ⁴Laboratory of Macromolecular Complexes, Center for Structural Biology of Challenging Proteins, The University of Tokyo, Tokyo 113-0032, Japan. ⁵Mitsubishi Kagaku Institute of Life Sciences, Machida-shi, Tokyo 194-8511, Japan. ⁶Department of Biochemistry and Molecular Genetics, Schools of Medicine and Dentistry, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA. ⁷Bioengineering Department, Nagaoka University of Technology, Niigata 940-2188, Japan. ⁸Kyoto Sangyo University, Kita-ku, Kyoto 603-8555, Japan.

*These authors contributed equally to this work.

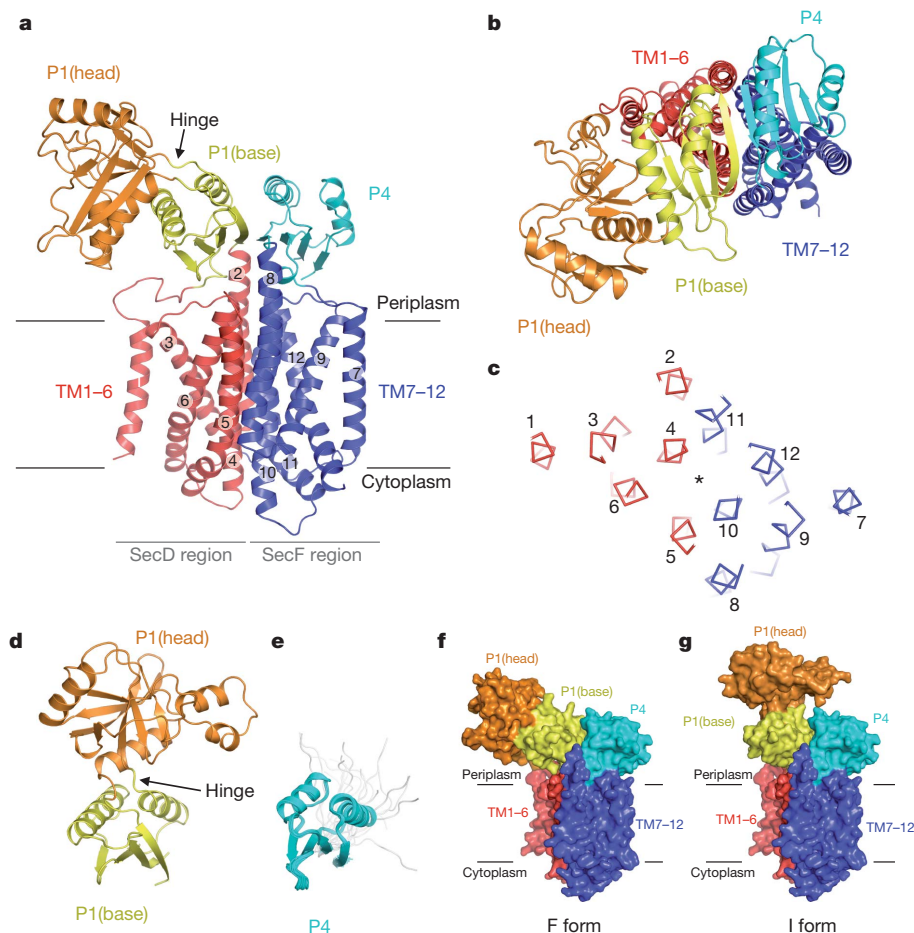


Figure 1 | Structures of *T. thermophilus* SecDF. **a, b**, The crystal structure of full-length SecDF viewed from the membrane side (**a**) and the periplasmic side (**b**). **c**, TtSecDF cross-sectioned at the middle of the TM, viewed from the periplasm. The asterisk indicates the pseudo-symmetrical axis. TMs are numbered. **d**, Crystal structure of the P1 domain. **e**, NMR structure of the P4 domain after twenty superimpositions. The disordered regions are shown in grey. **f**, Crystal structure of full-length SecDF, F form. **g**, Crystal structure of full-length SecDF, I form. The base subdomain of isolated P1 was docked onto that of the F form, as shown in (**f**).

SecDF variants (described below) did not support the PMF-dependent completion of pro-OmpA translocation (Fig. 2c).

To gain structural insights into the PMF-dependence of SecDF function, we compared its structure with that of AcrB²⁰, an RND superfamily proton/multi-drug antiporter¹⁶ (Supplementary Fig. 5). AcrB forms a homotrimer, whereas SecDF is monomeric. Although the TM segments of SecDF and AcrB share low sequence identity of 15% (Supplementary Fig. 6), the structures of their TM regions are similar, yielding a root mean square deviation of ~ 2.7 Å for the C α atoms of the TM helices. By contrast, the structures and functions of

the periplasmic regions are different between SecDF and AcrB. The TM region of AcrB is thought to participate in proton transport and contains several conserved, charged residues important for drug export activity, such as Asp 407, Asp 408, Lys 940 and Arg 971²¹ (Supplementary Fig. 5e). Asp 407, Thr 978 and Arg 971 in AcrB have structural counterparts in TtSecDF: Asp 340, Thr 675 and Arg 671 respectively (Fig. 3a and Supplementary Fig. 5e). The conserved SecDF residues are clustered at the TM interfaces between SecD and SecF and in the periplasmic base region underneath the head (Supplementary Fig. 7). We also note that Asp 637 is a highly conserved, membrane-embedded charged residue. Complementation tests indicated that the Asp519Asn mutation in EcSecD as well as the Asp213Asn and Arg247Met mutations in EcSecF (Fig. 3a) completely abolished SecDF activity and conferred some dominant-negative phenotypes (Fig. 3b). Thus, these conserved charged residues are crucial for SecDF function, consistent with the hypothesis that SecDF conducts protons through the conserved TM region.

The halophilic marine bacteria *Vibrio* spp. use a sodium ion (Na⁺) gradient instead of PMF for some cellular processes²². *Vibrio alginolyticus* has two sets of *secDF* genes encoding SecDF-1 and SecDF-2 complexes respectively. When VaSecDF-1 from *V. alginolyticus* 138-2 was expressed in the *E. coli* *secD1* (Cs) mutant, the addition of NaCl,

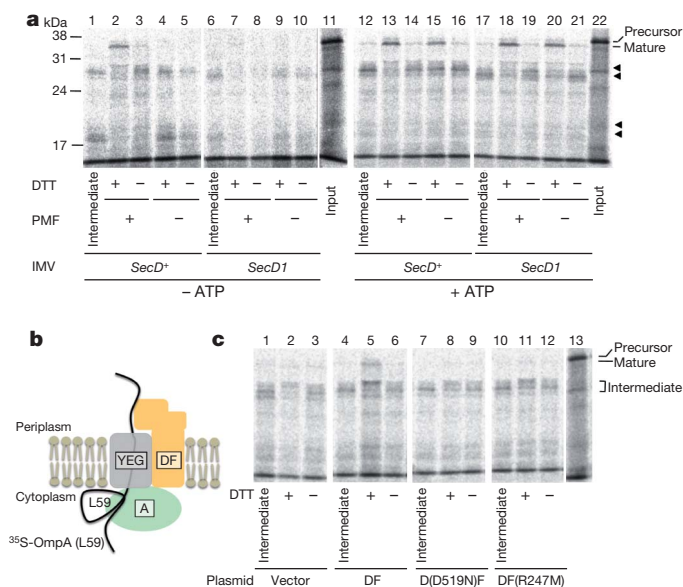


Figure 2 | SecDF-dependent translocation completion. **a**, Identification of a SecDF- and PMF-dependent translocation step. SecDF-deficient (*secD1*) IMVs were incubated with ³⁵S-labelled pro-OmpA(L59) to generate translocation intermediates. Protein translocation was continued in the presence or absence of ATP, PMF and dithiothreitol (DTT) before SDS-PAGE and phosphor imaging. Arrowheads indicate translocation intermediates. **b**, Schematic depiction of the translocation intermediate of pro-OmpA(L59). A, SecA; DF, SecDF; YEG, SecYEG. **c**, Completion of pro-OmpA(L59) translocation using IMVs from the *secD1* (Cs) mutant expressing no additional protein (vector), *E. coli* SecDF or the SecDF derivatives SecD(D519N)F and SecDF(R247M).

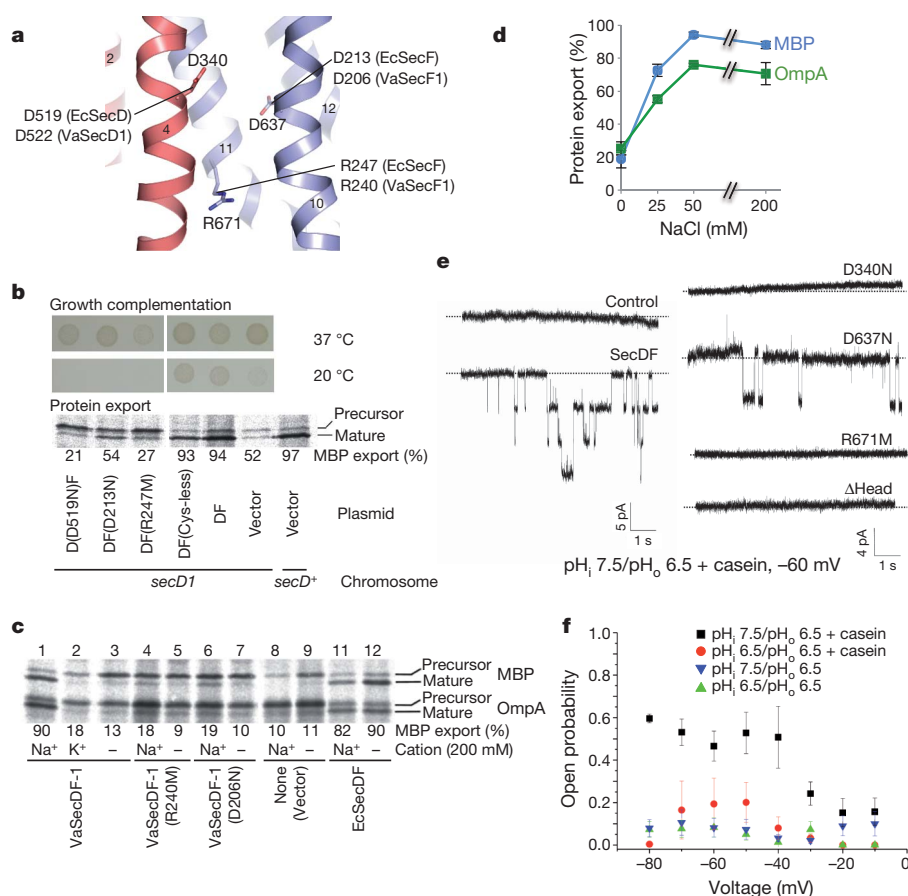


Figure 3 | Functional charged residues and proton conduction of SecDF. **a**, Functionally important, conserved residues in the TM regions. The TMs are numbered. **b**, Complementation of the *secD1* (Cs) growth and protein export defects by SecD and SecF mutants. Assay, SDS-PAGE and phosphor imaging as in Fig. 2. MBP, maltose-binding protein. **c**, Na⁺-dependent protein export by VaSecDF-1. **d**, Na⁺-dependence of protein export by VaSecDF-1 ($n = 3$). **e**, Single channel currents recorded by patch clamp in membrane patches excised from *E. coli* spheroplasts containing TtSecDF (left panel) or its mutants D340N, D637N, R671M and ΔHead (right panel) and effects of a pH gradient and casein. **f**, The channel open probability in experiments shown in (e) ($n = 6$ with casein, $n = 4$ without casein). Error bars indicate s.e.m.

but not KCl, to the medium restored protein export (Fig. 3c, lanes 1–3). Export was dependent on the extracellular Na⁺ concentration up to 50 mM (Fig. 3d) but such Na⁺-dependence was not observed with EcSecDF (Fig. 3c, lanes 11 and 12). Thus, VaSecDF-1 facilitates protein export using a Na⁺ gradient across the membrane. The Arg240Met and Asp206Asn mutations in VaSecF-1, corresponding to EcSecF Arg247Met and Asp213Asn respectively (Fig. 3a), compromised the Na⁺-dependent protein export activity (Fig. 3c, lanes 4–7). These results provide physiological evidence for cation-coupled protein translocation by SecDF.

To verify that SecDF utilizes the PMF and thus conducts protons, we performed inside-out patch clamp experiments²³ using *E. coli* giant spheroplasts²⁴ containing TtSecDF. Current recordings revealed that spheroplasts containing TtSecDF underwent transient channel openings under both symmetrical and asymmetrical pH conditions, whereas spheroplasts without TtSecDF did not (Supplementary Fig. 8a). The channel activities were markedly enhanced by the imposition of a pH gradient as well as by the addition of a P1-interacting unfolded protein, casein, to the pipette (periplasmic side) solution (Fig. 3e left panel, Supplementary Fig. 8a–c, Supplementary Discussion). In addition, casein and acidic extracellular conditions increased the probability of opening of the SecDF-dependent ion channel (Fig. 3f). Thus, ion-conduction can be regulated by a proton gradient and by the binding of an unfolded protein to SecDF. The use of a proton-specific fluorescent probe, BCECF-AM (2',7'-bis-(2-carboxyethyl)-5-(and-6)-carboxyfluorescein, acetoxymethyl ester), confirmed SecDF-dependent proton import (Supplementary Fig. 9 and Supplementary Discussion). The TtSecDF mutations Asp340Asn and Arg671Met, but not Asp637Asn, abolished ion channel activity (Fig. 3a and 3e right panel), indicating that Asp 340 and Arg 671 in the TM region are essential for proton conduction and might be protonated transiently. Asp 637, which is sequestered from the putative main proton pathway in the TM region of both AcrB and SecDF (Supplementary Fig. 5e), might function in subsequent conformational changes that are required for the enhancement of protein translocation.

In addition to the TM residues, the P1 head domain is a critical element for proton transport. TtSecDF(Δ112–248), which lacks the head, did not show any signals in the patch clamp assay (Fig. 3e right panel) and the corresponding EcSecDF mutant was defective in protein export (Supplementary Fig. 4f). The conformational flexibility of P1 is also required for proton conduction because the TtSecDF double cysteine mutant, Leu106Cys/Leu243Cys (see Supplementary Fig. 4a, b), lacked channel activity in four of the six membrane patches examined. The variable results could be explained by incomplete crosslinking (more than 20%, data not shown). We think it is likely that the putative hinge motion of the P1 domain is coupled with both proton transport and protein export. Taken together, we propose that proton flow is the driving force for P1 domain movement and for the consequent enhancement of protein translocation by SecDF.

We have shown here that SecDF and PMF are required for the post-initiation mode of translocation, which can occur in the absence of ATP and SecA. This function depends on the ability of the periplasmic P1 domain to interact with a substrate and to undergo a structural transition between the I and F forms, which are likely to be in equilibrium. The F state of SecDF may place the tilted P1 head above the translocon pore, enabling it to capture an emerging preprotein (Fig. 4a). The preprotein-bearing F form could then return to the I configuration (Fig. 4b), preventing the backward movement of the substrate and driving its forward movement. The release of the bound preprotein from SecDF and the subsequent I to F conversion may be coupled to proton flow (Fig. 4c). These action cycles will eventually lead to the completion of translocation, in which the substrate is released from the translocon. As the bacterial periplasm lacks ATP, SecDF may utilize the PMF to drive its conformational transition and delivery of substrates. The Asp and Arg residues in the TM region of SecDF could serve as putative proton acceptors in the proton relay pathway (Fig. 4c). The corresponding residues of AcrB (Supplementary Fig. 5e) have been proposed to participate in proton conduction with their side chains assuming different configurations among the asymmetric protomers,

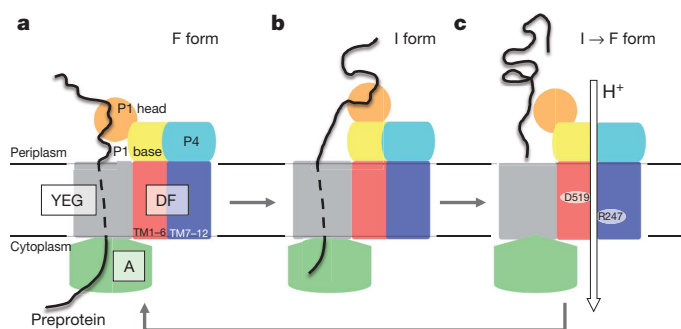


Figure 4 | A working model of the PMF-driven translocation enhancement by SecDF. **a**, F form, capturing state. **b**, I form, holding state. **c**, I to F transition and substrate-releasing state. The two essential charged residues of EcSecDF are highlighted. SecDF is coloured as in Fig. 2. SecYEG, grey; SecA, green; preprotein, black line; proton movement, white arrow. See the main text discussion for details.

which presumably exist in different protonation states²⁰. Likewise, protonation of the key charged residues of SecDF could induce the twisting of TM4 and TM10. This would be transmitted to the conserved P1-TM4 linker region (Supplementary Fig. 7) and would trigger the conformational transition of the P1 head subdomain.

Although direct evidence for the PMF-dependent conformational transition of SecDF will await further structural and functional studies, we have shown that SecDF is a component of the Sec machinery that utilizes the PMF to complete protein translocation after the ATP-dependent function of SecA.

METHODS SUMMARY

X-ray diffraction analysis of TtSecDF was described previously¹⁷. The initial phases were determined by the single-wavelength anomalous dispersion method. The initial model was manually built, using the structures of the TM regions of AcrB and the separately-determined P1 and P4 domain structures as references (Supplementary Methods). The model was substantially improved by using the zonal scaling²⁵ and methionine-marking²⁶ methods and was finally refined to $R_{\text{work}} = 29.8\%$ and $R_{\text{free}} = 31.9\%$ at 3.3 Å resolution.

To monitor the later steps of translocation, ³⁵S-labelled pro-OmpA(L59) and IMVs, prepared as described previously^{18,27}, were incubated to form translocation intermediates which were then incubated under several conditions and treated with proteinase K. The translocation status of ³⁵S-pro-OmpA was examined by phosphor imaging after SDS-polyacrylamide gel electrophoresis (SDS-PAGE).

For the complementation test, *E. coli secD1* (Cs) mutant cells carrying plasmids encoding HA-tagged EcSecD and EcSecF mutants were spotted onto LB agar plates (pH adjusted to ~7.5) and incubated at 20 °C. The efficiency of protein export *in vivo* was assessed by the ³⁵S-methionine pulse-labelling procedures.

To monitor the protein transport activity of *Vibrio* SecDF-1, *E. coli* secD1 cells expressing VaSecD-1 and VaSecF-1 were pulse-labelled in the presence or absence of Na⁺. To detect single channel activity, electric currents were measured with inside-out membrane patches excised from TtSecDF-containing *E. coli* giant spheroplasts, using an Axopatch 200B amplifier (Axon CNS, Molecular Devices).

Received 11 January 2010; accepted 9 March 2011.

Published online 11 May 2011.

- van den Berg, B. *et al.* X-ray structure of a protein-conducting channel. *Nature* **427**, 36–44 (2004).
- Zimmer, J., Nam, Y. & Rapoport, T. A. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455**, 936–943 (2008).
- Tsukazaki, T. *et al.* Conformational transition of Sec machinery inferred from bacterial SecY structures. *Nature* **455**, 988–991 (2008).
- du Plessis, D. J., Nouwen, N. & Driessen, A. J. The Sec translocase. *Biochim. Biophys. Acta* **1808**, 851–865 (2011).
- Driessen, A. J. & Wickner, W. Proton transfer is rate-limiting for translocation of precursor proteins by the *Escherichia coli* translocase. *Proc. Natl Acad. Sci. USA* **88**, 2471–2475 (1991).
- Shiozuka, K., Tani, K., Mizushima, S. & Tokuda, H. The proton motive force lowers the level of ATP required for the *in vitro* translocation of a secretory protein in *Escherichia coli*. *J. Biol. Chem.* **265**, 18843–18847 (1990).
- Pogliano, J. A. & Beckwith, J. SecD and SecF facilitate protein export in *Escherichia coli*. *EMBO J.* **13**, 554–561 (1994).
- Sagara, K., Matsuyama, S. & Mizushima, S. SecF stabilizes SecD and SecY, components of the protein translocation machinery of the *Escherichia coli* cytoplasmic membrane. *J. Bacteriol.* **176**, 4111–4116 (1994).

- Hand, N. J., Klein, R., Laskewitz, A. & Pohlschroder, M. Archaeal and bacterial SecD and SecF homologs exhibit striking structural and functional conservation. *J. Bacteriol.* **188**, 1251–1259 (2006).
- Matsuyama, S., Fujita, Y. & Mizushima, S. SecD is involved in the release of translocated secretory proteins from the cytoplasmic membrane of *Escherichia coli*. *EMBO J.* **12**, 265–270 (1993).
- Economou, A., Pogliano, J. A., Beckwith, J., Oliver, D. B. & Wickner, W. SecA membrane cycling at SecYEG is driven by distinct ATP binding and hydrolysis events and is regulated by SecD and SecF. *Cell* **83**, 1171–1181 (1995).
- Arkowitz, R. A. & Wickner, W. SecD and SecF are required for the proton electrochemical gradient stimulation of preprotein translocation. *EMBO J.* **13**, 954–963 (1994).
- Duong, F. & Wickner, W. The SecDFyajC domain of preprotein translocase controls preprotein movement by regulating SecA membrane cycling. *EMBO J.* **16**, 4871–4879 (1997).
- Nouwen, N., Piwowarek, M., Berrelkamp, G. & Driessen, A. J. The large first periplasmic loop of SecD and SecF plays an important role in SecDF functioning. *J. Bacteriol.* **187**, 5857–5860 (2005).
- Pogliano, K. J. & Beckwith, J. Genetic and molecular characterization of the *Escherichia coli* secD operon and its products. *J. Bacteriol.* **176**, 804–814 (1994).
- Tseng, T. T. *et al.* The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. *J. Mol. Microbiol. Biotechnol.* **1**, 107–125 (1999).
- Tsukazaki, T. *et al.* Purification, crystallization and preliminary X-ray diffraction of SecDF, a translocase-associated membrane protein, from *Thermus thermophilus*. *Acta Crystallogr. F* **62**, 376–380 (2006).
- Uchida, K., Mori, H. & Mizushima, S. Stepwise movement of preproteins in the process of translocation across the cytoplasmic membrane of *Escherichia coli*. *J. Biol. Chem.* **270**, 30862–30868 (1995).
- Schiebel, E., Driessen, A. J., Hartl, F. U. & Wickner, W. $\Delta\mu_{\text{H}^+}$ and ATP function at different steps of the catalytic cycle of preprotein translocase. *Cell* **64**, 927–939 (1991).
- Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T. & Yamaguchi, A. Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature* **443**, 173–179 (2006).
- Seeger, M. A., von Ballmoos, C., Verrey, F. & Pos, K. M. Crucial role of Asp408 in the proton translocation pathway of multidrug transporter AcrB: evidence from site-directed mutagenesis and carbodiimide labeling. *Biochemistry* **48**, 5801–5812 (2009).
- Häse, C. C. & Barquera, B. Role of sodium bioenergetics in *Vibrio cholerae*. *Biochim. Biophys. Acta* **1505**, 169–178 (2001).
- Sasaki, M., Takagi, M. & Okamura, Y. A voltage sensor-domain protein is a voltage-gated proton channel. *Science* **312**, 589–592 (2006).
- Hattori, M. *et al.* Mg²⁺-dependent gating of bacterial MgtE channel underlies Mg²⁺ homeostasis. *EMBO J.* **28**, 3602–3612 (2009).
- Vassilyev, D. G. *et al.* Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448**, 163–168 (2007).
- Inaba, K. *et al.* Crystal structure of the DsbB-DsbA complex reveals a mechanism of disulfide bond generation. *Cell* **127**, 789–801 (2006).
- Matsuo, E., Mori, H., Shimoike, T. & Ito, K. Syd, a SecY-interacting protein, excludes SecA from the SecY complex with an altered SecY24 subunit. *J. Biol. Chem.* **273**, 18835–18840 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Akiyama, R. Suno, Y. Morimoto, T. Minamino, K. Namba, K. Inaba, M. Hattori and H. Nishimatsu for suggestions; T. Sakamoto and A. Kurabayashi for assistance with sample preparation; R. Yamasaki, M. Sano, K. Mochizuki, K. Yoshikawa, K. Imaiyoishi and T. Adachi for technical support; M. Homma and S. Kojima for providing the *Vibrio* genomic DNA; the beamline staff members at BL41XU of SPring-8 (Hyogo, Japan) and at NW12 of KEK PF-AR (Tsukuba, Japan) for technical help during data collection and M. Iiba for comments on our manuscript. This work was supported by a Grant-in-Aid for Scientific Research (S) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to O.N., by a CREST grant from JST to K.I., by a BIRD grant from JST to H.M. and R.I., by a grant for the National Project on Protein Structural and Functional Analyses to O.N., by NIH grants to D.G.V. and by grants from MEXT to T. Tsukazaki, H.M., R.I., S.F. and K.I.

Author Contributions T. Tsukazaki performed the structural determination and the biochemical experiments with SecDF. H.M. performed the functional analyses of SecDF. Y.E. solved the crystal structure of the SecDF P1 domain and assisted with the functional analysis of SecDF. R.I., S.F., D.G.V. and O.N. assisted with the structural determination. A.D.M. performed patch clamp and pH fluorescence experiments. T. Tanaka and T.K. solved the structure of the P4 domain by NMR. A.P. and D.G.V. assisted with the crystallization and data collection of SecDF. All authors discussed the results and commented on the manuscript. O.N. and K.I. supervised the work and wrote and edited the manuscript.

Author Information The coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes 3AQP for the entire TtSecDF protein and 3AQO for the P1 domain. The PDB and BMRB codes for the deposited P4 domain are 2RRN and 11426 respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to O.N. (nureki@ims.u-tokyo.ac.jp) or K.I. (kito@cc.kyoto-su.ac.jp).

Non-apoptotic role of BID in inflammation and innate immunity

Garabet Yeretssian¹, Ricardo G. Correa², Karine Doiron¹, Patrick Fitzgerald³, Christopher P. Dillon³, Douglas R. Green³, John C. Reed² & Maya Saleh¹

Innate immunity is a fundamental defence response that depends on evolutionarily conserved pattern recognition receptors for sensing infections or danger signals^{1,2}. Nucleotide-binding and oligomerization domain (NOD) proteins are cytosolic pattern-recognition receptors of paramount importance in the intestine, and their dysregulation is associated with inflammatory bowel disease^{3,4}. They sense peptidoglycans from commensal microorganisms and pathogens and coordinate signalling events that culminate in the induction of inflammation and anti-microbial responses². However, the signalling mechanisms involved in this process are not fully understood. Here, using genome-wide RNA interference, we identify candidate genes that modulate the NOD1 inflammatory response in intestinal epithelial cells. Our results reveal a significant crosstalk between innate immunity and apoptosis and identify BID, a BCL2 family protein, as a critical component of the inflammatory response. Colonocytes depleted of BID or macrophages from *Bid*^{-/-} mice are markedly defective in cytokine production in response to NOD activation. Furthermore, *Bid*^{-/-} mice are unresponsive to local or systemic exposure to NOD agonists or their protective effect in experimental colitis. Mechanistically, BID interacts with NOD1, NOD2 and the I κ B kinase (IKK) complex, impacting NF- κ B and extracellular signal-regulated kinase (ERK) signalling. Our results define a novel role of BID in inflammation and immunity independent of its apoptotic function, furthering the mounting evidence of evolutionary conservation between the mechanisms of apoptosis and immunity.

To identify novel regulators of the NOD1 pathway, we conducted a small interfering RNA (siRNA) screen using a library targeting 7,170 human druggable genes. The human intestinal epithelial cell-line HT29 was used and interleukin (IL)-8 was measured as a readout⁵⁻⁷. A schematic of the screening strategy is presented in Fig. 1a. Our primary screen identified 227 and 198 genes that respectively enhanced or dampened IL-8 protein levels in response to the NOD1 agonist γ Tri-DAP (Fig. 1b and Supplementary Tables 1 and 2). To reduce potential off-target hits, we performed a validation screen by measuring IL-8 messenger RNA levels. A strong correlation between the primary and validation screens was observed (Fig. 1c), with 200 candidates retained independently of cell toxicity (Fig. 1c, d, Supplementary Fig. 1a, b and Supplementary Table 3). Next, we performed a counter-screen using tumour-necrosis factor (TNF) as an inducer of IL-8, because NOD1/2 and tumour necrosis factor receptor (TNFR1) pathways are structurally and functionally related^{8,9}. Altogether, 114 genes were identified as common candidate regulators of the NOD1/2 and TNFR1 pathways (Supplementary Fig. 1c and Supplementary Table 4) of which 60 were common to the NOD1 validated hits (Fig. 1d, Supplementary Fig. 1d and Supplementary Table 5).

To gain insights into the biological processes regulated by the uncovered candidates, we used PANTHER to cluster the primary hits according to their gene ontology (Supplementary Fig. 1e, f and Supplementary Tables 6–9). Genes involved in signal transduction,

metabolism and proliferation were significantly over-represented. Interestingly, genes of the apoptosis machinery were also significantly enriched. Analysis of the Novartis GNF human expression atlas data across 80 tissues showed that the candidate genes are predominantly expressed in immune and neuronal tissues (Wilcoxon $P < 0.05$; Supplementary Fig. 2a and Supplementary Tables 10–12). Furthermore, examination of their promoter region revealed significant enrichment in consensus sites for pro-inflammatory and mitogenic transcription factors (Supplementary Fig. 2b). These results are consistent with the established role of the NOD1 pathway in inflammation and IEC proliferation and support the validity of our screen. To define networks of interacting proteins, we used Ingenuity Pathway Analysis and manual data mining to construct an interaction network by anchoring on NOD1 pathway components. Network modelling identified functional groups linked to signal transduction, including known effectors of NOD1 signalling. Further analysis of this network identified the BH3-only BCL2 family member BID as a potential proximal regulator of the NOD1 signalling complex (Supplementary Fig. 2c).

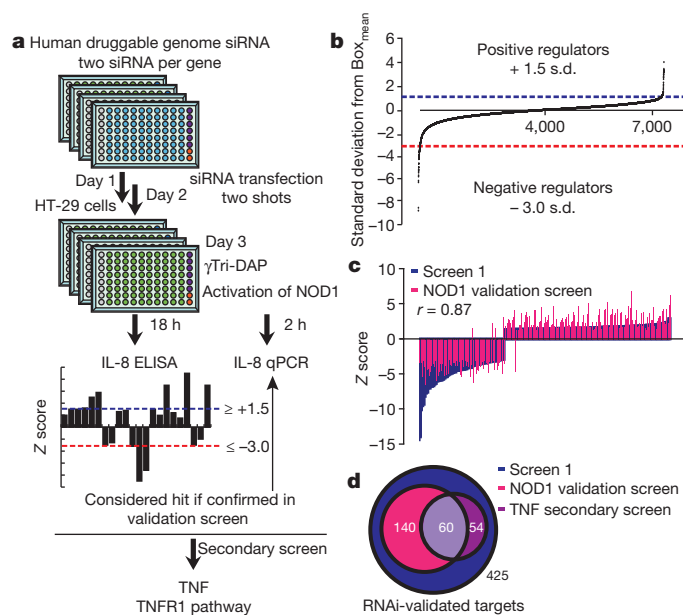


Figure 1 | Genome-wide RNA interference screen for genes regulating NOD1 signalling. **a**, Representation of the RNA interference screen strategy. **b**, Distribution of the data from the NOD1 primary screen. Each dot represents IL-8 concentration after individual gene depletion by a single siRNA pool. Data are represented as Z scores, with cut-offs (dashed lines) of 1.5 s.d. above the mean and 3.0 s.d. below the mean of all plates. **c**, Ranking by Z scores of 205 siRNAs from the primary and validation screens with a correlation coefficient $r = 0.87$. **d**, Venn diagram of hits from the primary, validation and secondary screens.

¹Department of Medicine, McGill University, Montreal, Quebec H3G 0B1, Canada. ²Sanford-Burnham Medical Research Institute, La Jolla, California 92037, USA. ³St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA.

Mounting evidence points to a co-evolution of the apoptosis and innate immunity pathways^{10–13}. Notably, NOD proteins and NOD-like receptors are structurally related to apoptosis-activating factor 1 (APAF1)^{10,11}, and the anti-apoptotic proteins BCL2 and BCL-X_L down-regulate innate immunity signalling by binding directly to NOD-like receptor proteins¹³. The interaction of BCL2 proteins with NOD-like receptors is reminiscent of that of CED9 with CED4, their respective orthologues in the *Caenorhabditis elegans* apoptosis pathway¹⁴.

To determine the role of BID in NOD1 signalling, we first validated the effect of BID depletion on γ Tri-DAP-induced IL-8 production in HT29 cells. Depletion of BID with three independent siRNAs led to a marked reduction in IL-8 production (Fig. 2a) similar to siRNAs against known positive regulators of the pathway (Fig. 2b and Supplementary Fig. 3a). Similarly, TNF and MCP-1 levels were dampened in BID-depleted cells in response to NOD1 activation (Supplementary Fig. 3a). Notably, depletion of other BCL2 family members including BCL2 or BAX did not modulate NOD signalling (Supplementary Fig. 3b). Consistent with the secondary screen results, BID was also required for TNF-induced IL-8 production (Supplementary Fig. 3c).

We next investigated the effect of BID depletion on the NOD1 signalling pathways. Figure 2c shows that BID impacted NF- κ B and ERK signalling. However, it was dispensable for p38 and Jun amino (N)-terminal kinases (JNK) activation (Supplementary Fig. 3d). The role of full-length BID is unclear and its function in apoptosis is contingent on its processing¹⁵. Furthermore, BID phosphorylation by

casein kinases I and II on serine residues proximal to the cleavage site hampers its processing and promotes cell survival¹⁶. Interestingly, γ Tri-DAP treatment of HT29 cells did not induce BID cleavage but resulted in its phosphorylation (Fig. 2c).

NOD1 is expressed in epithelial cells and fibroblasts, whereas NOD2 is highly expressed in myeloid cells¹⁷. We thus stimulated primary bone marrow-derived macrophages (BMDMs) from wild-type, *Ripk2*^{-/-} or *Bid*^{-/-} mice with the NOD2-specific agonist muramyl-dipeptide (MDP) as well as various Toll-like receptor (TLR) ligands and examined NOD2 signalling. IL-6 production was significantly blunted in BID-deficient macrophages in response to MDP alone or in synergy with lipopoly-saccharide (LPS), but not TLR ligands (Fig. 2d and Supplementary Fig. 3e, f). Similarly, MDP-stimulated production of TNF was dampened in *Bid*^{-/-} BMDMs compared with wild-type cells (Supplementary Fig. 3e). As in BID-depleted HT29 cells, *Bid*^{-/-} macrophages failed to activate NF- κ B and ERK signalling in response to NOD stimulation (Fig. 2e), whereas JNK and p38 activation was normal (Supplementary Fig. 3d). The role of BID in signal transduction in intestinal epithelial cells and macrophages was independent of apoptosis induction by NOD1 or NOD2 agonists (Supplementary Fig. 3g). These results are consistent with recent findings that identified a role of BID in the activation of the NF- κ B pathway and cell survival¹⁸.

To explore the role of BID processing and its BH3 domain in inflammation, we reconstituted *Bid*^{-/-} BMDMs with either wild-type BID, a non-cleavable form of BID (BID_{D59E}) or an apoptosis-deficient

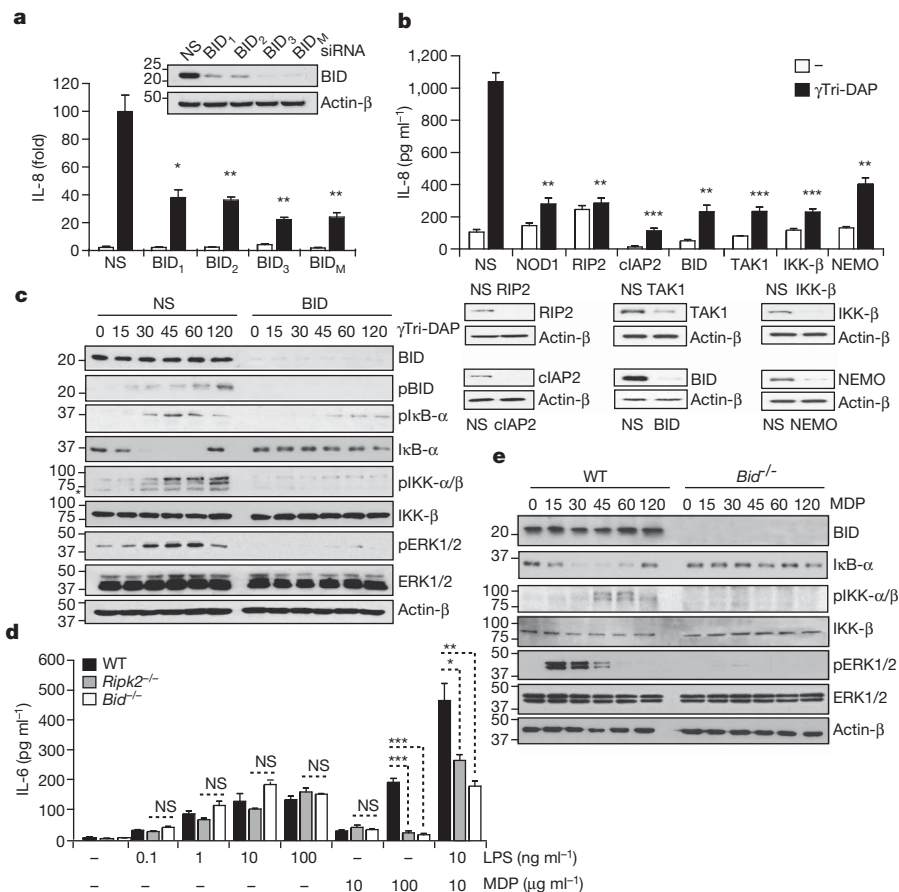


Figure 2 | BID is required for NOD1 and NOD2 signalling. **a**, HT29 cells were transfected with non-specific (NS) or three BID siRNA (BID₁, BID₂ or BID₃) alone or mixed (BID_M). IL-8 levels were determined by qPCR 2 h after DAP treatment. **b**, HT29 cells were treated as in **a**. IL-8 production was quantified by ELISA 18 h after DAP treatment. Protein depletion was assessed by western blotting (insets in **a** and **b**). **c**, HT29 cells were treated as in **a** for the indicated time points. I κ B α degradation and phosphorylation and MAPK, IKK α / β and BID phosphorylation were assayed by immunoblotting.

d, BMDMs from wild-type (WT, $n = 4$), *Ripk2*^{-/-} ($n = 4$) and *Bid*^{-/-} ($n = 4$) mice were left untreated or stimulated with LPS, MDP or MDP with LPS for 18 h. IL-6 production was quantified by ELISA. **e**, BMDMs from wild-type and *Bid*^{-/-} mice were treated with MDP. I κ B α degradation and MAPK, IKK α / β and BID phosphorylation were assayed by immunoblotting. p, phosphorylated. NS, not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. Values in graphs represent mean \pm s.e.m. from three or four independent experiments.

mutant form of BID (BID_{G94E})¹⁹, and examined cytokine production induced by MDP. Figure 3a shows that BID, BID_{D59E} and BID_{G94E} were equivalent in rescuing the blunted inflammatory phenotype of $Bid^{-/-}$ cells. These results suggested that the role of BID in inflammation is independent of its processing, its BH3 domain or its association with BAX. This was further confirmed in $Bax^{-/-}Bak^{-/-}$ macrophages from tamoxifen-treated $ROSA-CreER Bax^{fl/-}Bak^{-/-}$ mice that showed an unaltered response to MDP compared with wild-type cells (Fig. 3b). In contrast, mutation of reported phosphorylation sites in BID (BID_{S64A} , BID_{S65A} and BID_{S78A}) significantly blunted MDP-induced IL-6 production (Fig. 3a).

To examine whether BID associated with components of the NOD signalosome, we performed co-immunoprecipitation experiments and showed that BID interacted with NOD1 and NOD2 (Supplementary Fig. 4a). Similar results were obtained with BID_{G94E} (Supplementary Fig. 4b), but not with the serine to alanine mutants of BID, which were deficient in NOD1 binding (Supplementary Fig. 4c). *In vitro* pull-down experiments indicated that binding of NOD1 and NOD2 to purified recombinant BID was mediated through the central region encompassing the nucleotide-binding domain (Supplementary Fig. 4d). This result

is comparable to the binding of BCL2/BCL-X_L to NLRP1 (ref. 13). The region in BCL2/BCL-X_L required for this association has been previously mapped to an unstructured loop between α -helices 1 and 2 (ref. 20). Interestingly, full-length BID contains such a loop (residues 40–80)²¹. BID did not interact with NLRP1 (Supplementary Fig. 4e); however, truncation of its loop markedly reduced its association with NOD1 and NOD2 (Supplementary Fig. 4f). Consistently, subtle perturbation of the helical structure around the loop, through mutagenesis of specific leucine residues (L37A, L40A and L44A), drastically weakened the BID–NOD1 interaction (Supplementary Fig. 4g).

Upon agonist sensing, NOD1 and NOD2 associate with RIP2 and TAK1, which leads to activation of NF- κ B and MAPK (ERK1/2, p38 and JNK) pathways². The bifurcation in signalling between the NF- κ B and ERK versus p38 and JNK pathways occurs downstream of TAK1 and might involve differential recruitment of adaptor proteins that direct signalling towards specific inflammatory responses. TAK1 binds to IKK- γ (NEMO), MKK3/6 and MKK4/7 to trigger NF- κ B, p38 and JNK activation, respectively; whereas signalling to ERK1/2 occurs through an interaction between IKK- β and tumour progression locus 2 (TPL2)^{22–24}. Notably, the adaptor protein CARD9 transduces NOD signalling to p38 and JNK but not the NF- κ B pathway²⁵. The distinct requirement of BID in NF- κ B and ERK activation points to a possible interaction of BID with a signalling protein that associates with the IKK complex. Co-immunoprecipitation experiments revealed that BID associates with protein complexes containing IKK α , IKK- β and NEMO (Supplementary Fig. 4h). Reciprocally, NEMO associated with protein complexes containing BID and RIP2 (Supplementary Fig. 4i). To support these results further, we examined the association of endogenous BID with the NOD complex. Figure 3c shows that endogenous BID interacted with RIP2 in an agonist-dependent manner. Additionally, recruitment of the IKK complex to the NOD signalosome was mediated by BID, as NEMO interacted with RIP2 in wild-type but not $Ripk2^{-/-}$ or $Bid^{-/-}$ BMDMs after MDP stimulation (Fig. 3c). Immunoprecipitation of endogenous NEMO also revealed interaction with BID, IKK- β and RIP2-containing complexes in response to NOD activation (Fig. 3c). Altogether, these biochemical findings suggest that BID bridges the NOD receptors to the IKK complex (Fig. 3d).

To examine the function of BID in inflammation and innate immunity *in vivo*, we assessed the systemic and local responses of wild-type, $Ripk2^{-/-}$ and $Bid^{-/-}$ mice to administration of NOD agonists. Consistent with previous reports^{26,27}, intraperitoneal challenge with γ Tri-DAP induced a robust systemic production of chemokines and cytokines (Fig. 4a) and a pronounced peritonitis marked by infiltration of Gr1⁺ neutrophils (Fig. 4b) in wild-type mice but not $Ripk2^{-/-}$ or $Bid^{-/-}$ mice (Fig. 4a, b). Similarly, IL-6 production was dampened in $Ripk2^{-/-}$ and $Bid^{-/-}$ mice compared with wild-type mice in response to MDP administration (Supplementary Fig. 5a) but not that of LPS (Supplementary Fig. 5b). It has been previously shown that administration of the TLR3 ligand poly I:C and the NOD2 agonist MDP protect mice from dextran-sulphate sodium (DSS)-induced colitis^{26,28}. The protective effect of MDP is abrogated in mice deficient in NOD2, RIP2 or cIAP2 (refs 26, 27). To explore the role of BID in the NOD2 pathway, we subjected wild-type, $Ripk2^{-/-}$ and $Bid^{-/-}$ mice to an acute DSS treatment. Apart from wild-type mice treated with MDP, all other mice showed severe weight loss (Fig. 4c) and thickening and shortening of the colons upon necropsy (Supplementary Fig. 5d, e). Histological examination of colon sections further supported these data, as wild-type mice treated with MDP and DSS showed evidence of restored colonic crypt architecture, whereas all other mice exhibited severe leukocyte recruitment and crypt loss (Supplementary Fig. 5f). In contrast, $Ripk2^{-/-}$ and $Bid^{-/-}$ mice reacted similarly to wild-type mice in response to poly I:C (Supplementary Fig. 5c–f). Thus loss of BID function impaired the ability of NOD2 but not TLR3 to trigger tissue repair responses upon injury with DSS.

In summary, our high-throughput functional screen has identified approximately 200 candidate modulators of the NOD1 pathway and

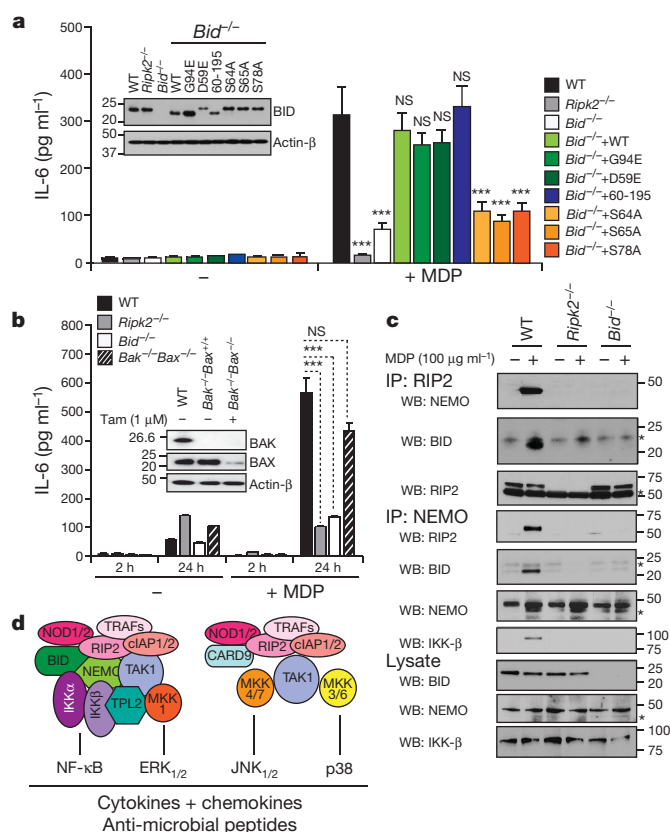


Figure 3 | BID regulates NOD signalling independently of its apoptotic determinants by linking NOD proteins to the IKK complex. **a**, BMDMs from $Bid^{-/-}$ mice ($n = 4$) were reconstituted with control or full-length BID, BID_{G94E} , BID_{D59E} , $BID_{(60-195)}$, BID_{S64A} , BID_{S65A} or BID_{S78A} mutants and stimulated with MDP for 18 h. IL-6 production was quantified by ELISA. BID expression was determined by immunoblotting (inset). **b**, BMDMs were isolated from wild-type ($n = 4$), $Ripk2^{-/-}$ ($n = 4$), $Bid^{-/-}$ ($n = 4$) and $ROSA-CreER Bax^{fl/-}Bak^{-/-}$ ($n = 3$) mice. 4-OH tamoxifen was used to excise the *Bax* gene. BMDMs were challenged with MDP for the indicated time points and IL-6 production was measured by ELISA. The efficiency of *Bax* excision was determined by immunoblotting (inset). **c**, BMDMs from wild-type, $Ripk2^{-/-}$ and $Bid^{-/-}$ mice were left untreated or treated with MDP for 30 min. Cell lysates were used for immunoprecipitation using anti-RIP2 or NEMO antibodies and analysed by immunoblotting. NS, not significant; *** $P < 0.001$. Asterisk, IgG heavy chain. Values in graphs represent mean \pm s.e.m. from three independent experiments. **d**, Representation of the NOD signalosome.

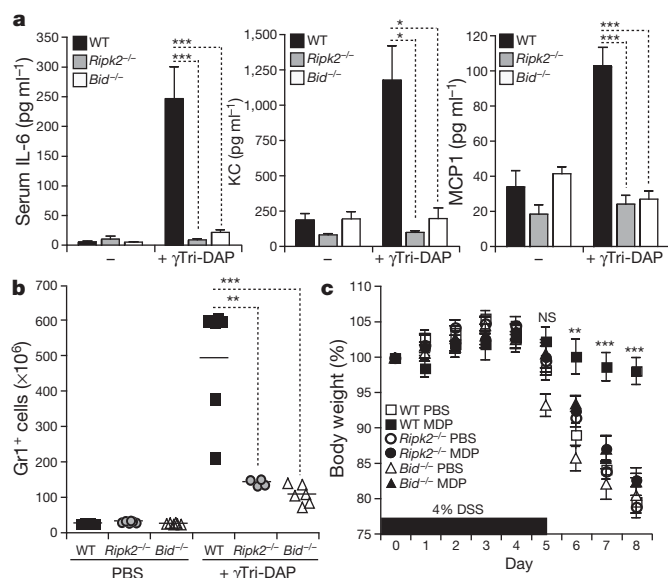


Figure 4 | *Bid*^{-/-} mice exhibit a blunted inflammatory response after NOD stimulation *in vivo* and are not protected from DSS colitis by MDP. **a**, Wild-type (WT, *n* = 6), *Ripk2*^{-/-} (*n* = 4) and *Bid*^{-/-} (*n* = 6) mice were injected intraperitoneally with γTri-DAP or endotoxin-free PBS for 2 h. IL-6, KC and MCP-1 levels in the serum were measured by ELISA. **b**, Mice were injected intraperitoneally with γTri-DAP for 24 h. The number of infiltrating Gr1⁺ cells in the peritoneum was quantified by fluorescence-activated cell sorting. **c**, Wild-type (*n* = 8), *Ripk2*^{-/-} (*n* = 8) and *Bid*^{-/-} (*n* = 8) mice were treated with 4% DSS. On days 0, 1 and 2, mice were injected intraperitoneally with MDP or endotoxin-free PBS. Body weight loss is represented. NS, not significant; **P* < 0.05; ***P* < 0.01; ****P* < 0.001. Values in graphs represent mean ± s.e.m. from three independent experiments.

provides a comprehensive view of the biological processes involved in NOD signalling. Importantly, we have uncovered the physiological function of full-length BID as a critical mediator of inflammation and innate immunity. Our results place BID at the centre of the cell's decision to induce innate immune responses (and survive) or commit suicide by apoptosis. We anticipate that the development of small molecules targeting BID will set the stage for the emergence of novel strategies to treat immune-mediated inflammatory diseases including inflammatory bowel disease.

METHODS SUMMARY

The siRNA screen was performed using the human druggable genome siRNA set V2.0 library from Qiagen. HT29 cells were transfected with siRNA (50 nM) using Lipofectamine 2000 and treated with DOTAP conjugated γTri-DAP (10 μg ml⁻¹). Cytokines and chemokines were measured by enzyme-linked immunosorbent assay (ELISA) and quantitative real-time PCR (qPCR). Data were analysed using *cellHTS2*. Bioinformatics were analysed using PANTHER, MSigDB and TRANSFAC, and network modelling was performed with data obtained from the Human Protein Reference Database, the Biomolecular Interaction Network Database and the Ingenuity Pathways Analysis database. BMDMs were treated with DOTAP-conjugated MDP (10–100 μg ml⁻¹). Reconstitution of BMDMs was performed with Amaxa (Solution T, program T-20). NOD signalling was assessed by immunoblotting using total and phospho-specific antibodies. Co-immunoprecipitation experiments were performed in HEK293T cells. γTri-DAP (1 mg kg⁻¹), MDP (100 μg) and poly I:C (100 μg) were injected intraperitoneally in mice, and 4% DSS was administered in the drinking water.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 August 2010; accepted 3 March 2011.

Published online 8 May 2011.

- Garrett, W. S., Gordon, J. I. & Glimcher, L. H. Homeostasis and inflammation in the intestine. *Cell* **140**, 859–870 (2010).

- Takeuchi, O. & Akira, S. Pattern recognition receptors and inflammation. *Cell* **140**, 805–820 (2010).
- Cho, J. H. The genetics and immunopathogenesis of inflammatory bowel disease. *Nature Rev. Immunol.* **8**, 458–466 (2008).
- Turner, J. R. Intestinal mucosal barrier function in health and disease. *Nature Rev. Immunol.* **9**, 799–809 (2009).
- Girardin, S. E. *et al.* Nod1 detects a unique muropeptide from Gram-negative bacterial peptidoglycan. *Science* **300**, 1584–1587 (2003).
- Kim, J. G., Lee, S. J. & Kagnoff, M. F. Nod1 is an essential signal transducer in intestinal epithelial cells infected with bacteria that avoid recognition by toll-like receptors. *Infect. Immun.* **72**, 1487–1495 (2004).
- Masumoto, J. *et al.* Nod1 acts as an intracellular receptor to stimulate chemokine production and neutrophil recruitment *in vivo*. *J. Exp. Med.* **203**, 203–213 (2006).
- Chen, C. M., Gong, Y., Zhang, M. & Chen, J. J. Reciprocal cross-talk between Nod2 and TAK1 signaling pathways. *J. Biol. Chem.* **279**, 25876–25882 (2004).
- Kobayashi, K. S. *et al.* Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* **307**, 731–734 (2005).
- Bertin, J. *et al.* Human CARD4 protein is a novel CED-4/Apaf-1 cell death family member that activates NF-κB. *J. Biol. Chem.* **274**, 12955–12958 (1999).
- Inohara, N. *et al.* Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-κB. *J. Biol. Chem.* **274**, 14560–14567 (1999).
- Faustin, B. *et al.* Reconstituted NALP1 inflammasome reveals two-step mechanism of caspase-1 activation. *Mol. Cell* **25**, 713–724 (2007).
- Bruey, J. M. *et al.* Bcl-2 and Bcl-XL regulate proinflammatory caspase-1 activation by interaction with NALP1. *Cell* **129**, 45–56 (2007).
- Conradt, B. & Horvitz, H. R. The *C. elegans* protein EGL-1 is required for programmed cell death and interacts with the Bcl-2-like protein CED-9. *Cell* **93**, 519–529 (1998).
- Strasser, A. The role of BH3-only proteins in the immune system. *Nature Rev. Immunol.* **5**, 189–200 (2005).
- Desagher, S. *et al.* Phosphorylation of bid by casein kinases I and II regulates its cleavage by caspase 8. *Mol. Cell* **8**, 601–611 (2001).
- Xavier, R. J. & Podolsky, D. K. Unravelling the pathogenesis of inflammatory bowel disease. *Nature* **448**, 427–434 (2007).
- Luo, W. *et al.* Bid mediates anti-apoptotic COX-2 induction through the IKKβ/NFκB pathway due to 5-MCDE exposure. *Curr. Cancer Drug Targets* **10**, 96–106 (2010).
- Wang, K., Yin, X. M., Chao, D. T., Milliman, C. L. & Korsmeyer, S. J. BID: a novel BH3 domain-only death agonist. *Genes Dev.* **10**, 2859–2869 (1996).
- Faustin, B. *et al.* Mechanism of Bcl-2 and Bcl-X(L) inhibition of NLRP1 inflammasome: loop domain-dependent suppression of ATP binding and oligomerization. *Proc. Natl Acad. Sci. USA* **106**, 3935–3940 (2009).
- Petros, A. M., Olejniczak, E. T. & Fesik, S. W. Structural biology of the Bcl-2 family of proteins. *Biochim. Biophys. Acta* **1644**, 83–94 (2004).
- Hasegawa, M. *et al.* A critical role of RICK/RIP2 polyubiquitination in Nod-induced NF-κB activation. *EMBO J.* **27**, 373–383 (2007).
- Abbott, D. W., Wilkins, A., Asara, J. M. & Cantley, L. C. The Crohn's disease protein, NOD2, requires RIP2 in order to induce ubiquitylation of a novel site on NEMO. *Curr. Biol.* **14**, 2217–2227 (2004).
- Perkins, N. D. Integrating cell-signalling pathways with NF-κB and IKK function. *Nature Rev. Mol. Cell Biol.* **8**, 49–62 (2007).
- Hsu, Y. M. *et al.* The adaptor protein CARD9 is required for innate immune responses to intracellular pathogens. *Nature Immunol.* **8**, 198–205 (2007).
- Watanabe, T. *et al.* Muramyl dipeptide activation of nucleotide-binding oligomerization domain 2 protects mice from experimental colitis. *J. Clin. Invest.* **118**, 545–559 (2008).
- Bertrand, M. J. *et al.* Cellular inhibitors of apoptosis cIAP1 and cIAP2 are required for innate immunity signaling by the pattern recognition receptors NOD1 and NOD2. *Immunity* **30**, 789–801 (2009).
- Vijay-Kumar, M. *et al.* Activation of toll-like receptor 3 protects against DSS-induced acute colitis. *Inflamm. Bowel Dis.* **13**, 856–864 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Strasser for providing *Bid*^{-/-} mice, G. Shore for BID antibodies and the McGill University high throughput/high content screening facility. We also thank D. Zhai for BID purification. This work was supported by grants from the Canadian Institutes for Health Research (CIHR-MOP 82801) and the Burroughs Wellcome Fund to M.S. M.S. is a Canadian Institutes for Health Research New Investigator. G.Y. is supported by a PDF-Fellowship from the McGill University Health Center. C.P.D. is supported by a fellowship grant from the SASS Foundation for Medical Research.

Author Contributions G.Y. and M.S. designed the research. G.Y., K.D. and M.S. performed the screen. G.Y. and M.S. analysed the data. G.Y. performed most experiments. R.G.C. performed Ni/NTA pull-down assay; P.F., R.G.C., C.P.D., D.R.G. and J.C.R. contributed new reagents/analytical tools. G.Y. and M.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.S. (maya.saleh@mcgill.ca).

METHODS

Animal strains. Wild-type (C57Bl/6) *Bid*^{-/-} mice from A. Strasser and *Ripk2*^{-/-} from R. Flavell were backcrossed to a pure C57Bl/6 genetic background. Housing and all animal procedures were approved by McGill University operating under the guidelines of the Canadian Council of Animal Care.

Reagents and plasmids. Antibodies for pJNK (9255), pp38 (9211), p38 (9212), pERK1/2 (9106S), ERK1/2 (9102), IκBα (4812), pIκBα (9246 s), IKK-β (2370) and pIKK-α/β (2697) were purchased from Cell Signaling Technologies. Anti-actin-β (A1978) was from Sigma; anti-JNK1/3 (sc-474), anti-RIP2 (sc-22763), anti-BID (sc-11423), anti-BAX (sc-493), anti-NEMO (sc-8330) and anti-BAK (sc-832) were from Santa Cruz Biotechnology. Monoclonal anti-RIP2 was obtained from BD Bioscience (612349), anti-TAK1 (7263) from Millipore and cIAP2-specific antibody from R&D systems (AF8181). Antibody against pBid-S61 (A300-527A) was obtained from Bethyl laboratories. The Flag antibody was purchased from Sigma (F3165). All other antibodies against tags anti-haemagglutinin (11583816001), anti-Myc (11667149001) and anti-green fluorescent protein (11814460001) were obtained from Roche. Secondary antibodies were from Jackson ImmunoResearch Laboratories. All siRNAs used in this study were purchased from Qiagen. γTri-DAP (60774) was from Anaspec, MDP (G-1055) was from Bachem, and human recombinant TNF-α (30001A) and mouse recombinant TNF-α (31501) were from Peprotech. Cyclohexamide (C4859) was from Sigma and AnnexinV-FITC (556420) was from BD Pharmingen. DOTAP transfection reagent (11202375001) was from Roche, and DSS (relative molecular mass 36,000–50,000), was from MP Biomedicals (160110).

To generate the non-cleavable form of BID (BID_{D59E}), the apoptosis-deficient mutant form of BID (BID_{G94E}), the serine to alanine mutant forms of BID (BID_{S64A}, BID_{S65A} and BID_{S78A}) as well as the leucine to alanine mutant forms of BID (BID_{L37A}, BID_{L40A} and BID_{L44A}) mutagenesis PCRs on pcDNA3.1-HA-BID or pEGFP-BID were performed using a site-directed mutagenesis kit from Stratagene (200519-5). HA-tagged BID deletion constructs were obtained by PCR sub-cloning into pcDNA3.1 with the primers listed in Supplementary Table 13. The plasmids expressing Flag-tagged IKK-α and IKK-β were obtained from Tularik. The Flag-tagged NOD2 construct was obtained by inserting a Flag sequence in frame with the human NOD2 sequence in a pcDNA3 plasmid originally from G. Nunez. Myc-tagged NLRP1 was described previously^{9,16}. Myc-tagged NEMO was obtained from I. Verma. The construct expressing Flag-tagged RIP2 was a gift of G. Nunez; Flag-NOD1, RIP2-CARD and RIP2-ΔCARD plasmids were described previously²⁵. The plasmid encoding HA-TAK1 was provided by S. Baksh. Myc-tagged NOD1 and NOD2, as well as their deletion constructs, were described previously²⁹.

Cell culture. HEK293T and HT29 cells were maintained in 5% CO₂ at 37 °C in Dulbecco's modified Eagle's medium (HyClone) and McCoy's 5A modified medium (HyClone), respectively, supplemented with 10% fetal calf serum (HyClone), 2 mM L-glutamine and 100 μg ml⁻¹ penicillin/streptomycin. siRNA as well as plasmids were transfected with Lipofectamine 2000 (Invitrogen, 11668-19). We treated HT29 cells for different time points (0–120 min) with γTri-DAP at 10 μg ml⁻¹ transfected into the cells by using the DOTAP cationic lipid transfection reagent in accordance with the manufacturer's instructions. Cell-culture supernatants were used to quantify IL-8 levels by ELISA. For cell-death assay, HT29 cells were treated with γTri-DAP (10 μg ml⁻¹) or TNF-α (10 ng ml⁻¹) alone or in combination with cyclohexamide (10 μg ml⁻¹) for 24 h and 48 h.

siRNA screen and data analysis. siRNA screen was performed using the human druggable genome siRNA set V2.0 library from Qiagen and the Biomek II liquid handler from Beckman Coulter. The library consists of a pool of two siRNA duplexes for each target gene. A total of 7,170 human genes were assayed in the primary screen. For the validation screen, siRNAs for each gene used in the first screen were procured separately and tested. The library was received in 89-well × 96-well plates at 5 μM stock concentration. We prepared daughter plates at 1 μM that we used in the screen. HT29 cells (5 × 10³ per well) were transfected with siRNA (50 nM) using Lipofectamine 2000. After 24 h a second siRNA transfection was performed to ensure high efficiency of knockdown, and plates were incubated for an additional 24 h before treatment with DOTAP-encapsulated γTri-DAP (10 μg ml⁻¹) or recombinant human TNF-α (10 ng ml⁻¹). For the primary and secondary screens, cells were treated for 16 h with γTri-DAP or TNF-α and culture supernatants collected to measure IL-8 production by ELISA. For the validation screen, cells were treated for 2 h with γTri-DAP and IL-8 messenger RNA levels were quantified by qPCR. The raw data obtained from the primary screen were analysed using the open source Bioconductor/R package, *cellHTS2* (<http://www.dkfz.de/signaling/cellHTS>)³⁰. This program generated annotated gene lists with the corresponding adjusted Z scores (s.d. from the mean value of the control wells). We defined hits as positive or negative regulators according to a Z-score cut-off. Our cut-off criterion was 1.5 s.d. above the mean value for positive regulators and 3.0 s.d. below the mean for negative regulators. To evaluate the robustness and sensitivity of our screen, we calculated the overall Z'

factor of the screen as previously described³¹ taking into account the repeated positive and negative controls. The Z' factor obtained was greater than 0.5, confirming the overall quality of our assay.

Enrichment analyses. Genes were classified into biological process and molecular function categories using the PANTHER classification system³² and transcription-factor-binding sites using MSigDB and TRANSFAC v7.4 (www.gene-regulation.com). To assess the statistical enrichment or overrepresentation of these categories for the hit genes relative to their representation in the global set of genes examined in the siRNA screen, P values were computed using the hypergeometric test³³, which was implemented in the R language (<http://www.r-project.org/>). Briefly, the hypergeometric distribution describes the probability of finding at least s genes associated with a particular category in a set of g genes involved in NOD1 innate immune response, given that there are S genes associated with that same category in the global set of G genes examined in the genome-wide siRNA screen. For each category, c, and the list of genes l, the P value was calculated as $P(c, l) = 1 - \sum_{k=0}^{c-l} \frac{C(c, k)C(G-g, S-k)}{C(G, S)}$. The binomial coefficient is of the form $C(n, r)$. A value of P < 0.05 was considered significant. Categories assigned with at least ten genes are displayed in Supplementary Figs 2 and 3b.

Human protein interaction networks. The protein network was constructed by iteratively connecting interacting proteins, with data obtained from the Human Protein Reference Database, the Biomolecular Interaction Network Database, the Ingenuity Pathways database and functional information from the literature. The network uses graph theoretical representation in which components (gene products) are depicted as nodes and interactions between components as edges, implemented in the Perl programming language.

Analysis of gene expression across 80 tissues. Microarray data files were obtained from the Novartis GNF human expression atlas version-2 resource. Expression values of 33,689 probe sets from the HG-U133A (Affymetrix) platform and the GNF1H custom chip were analysed. The data set was normalized using global median scaling, and we filtered the data by excluding from the analysis those probe sets with 100% 'absent' calls (MAS5.0 algorithm) across all 80 tissues. The data set was further filtered by setting a minimum threshold value greater than 20 in at least one sample for each probe set and a maximum-mean expression value greater than 100. Hierarchical clustering (centroid linkage method) was performed with Cluster 3.0 using Pearson's correlation as the similarity metric³⁴. Z-score transformation was applied to each probe set across all arrays before generating 'heat maps' for visualization using TreeView³⁵.

Primary culture, transfection and treatment of bone-marrow-derived macrophages. Femurs and tibias of wild-type (C57Bl/6), *Ripk2*^{-/-}, *Bid*^{-/-} and *ROSA-CreER Bax*^{fl/fl} *Bak*^{-/-} mice were flushed with a 25G needle and cold RPMI 1640 (GIBCO) into a 15 ml tube. The suspension was then filtered through a 70 μm cell strainer and spun for 5 min at 400g. Pellets were re-suspended in complete RPMI 1640 supplemented with 15% L929 cell-conditioned media and plated in non-cell-culture-treated 10 cm dishes. Cells were cultured at 37 °C in a 10% CO₂ tissue culture incubator for 5–6 days, with media changed on day 3. MDP (at 10 or 100 μg ml⁻¹) was transfected into cells for different time points (0–120 min) with DOTAP reagent according to the manufacturer's instructions. BMDMs were transfected with BID wild-type and BID_{G94E} plasmids using the AMAXA Nucleofector (Solution T, program T-20; AMAXA biosystems) and treated with MDP (100 μg ml⁻¹). *ROSA-CreER Bax*^{fl/fl} *Bak*^{-/-} cells were subjected to 1 μM of 4-OH-tamoxifen (Sigma, T5648) during the differentiation period, followed by MDP treatment for 18 h. BMDMs were treated for 18 h with various TLR ligands: PAMcsk4 (1 μg ml⁻¹, TLR2), poly I:C (25 μg ml⁻¹, TLR3), LPS (*Escherichia coli* O111:B4; 0.1–100 ng ml⁻¹ and 1 μg ml⁻¹, TLR4), PGN (*Bacillus subtilis*; 2 μg ml⁻¹, flagellin (2 μg ml⁻¹, TLR5), resiquimod R848 (10 μg ml⁻¹, TLR7) and CpG (1 μg ml⁻¹, TLR9). In all these experiments, IL-6 production was measured in cell-culture supernatants using ELISA. For cell death assay, BMDMs were treated with MDP (100 μg ml⁻¹) or TNF-α (10 ng ml⁻¹) alone or in combination with cyclohexamide (10 μg ml⁻¹) for 24 h.

ELISA and lactate dehydrogenase assay. Cytokine and chemokine concentrations in cell-culture supernatants as well as sera were measured by ELISA using human IL-8/CXCL8 (DY208) and mouse IL-6 (DY406), KC/CXCL1 (DY453) and MCP-1/CC12 (DY479) kits purchased from R&D systems, used according to the manufacturer's instructions. For lactate dehydrogenase measurement, the CytoTox-96 non-radioactive cytotoxicity assay (Promega) was used according to the manufacturer's recommendations. The percentage of lactate dehydrogenase release was calculated compared with 100% cell lysis control.

Annexin V-FITC and propidium iodide staining. HT29 cells or BMDMs treated as described above were collected and re-suspended in 0.5 ml of Annexin V staining buffer (10 mM HEPES, 150 mM NaCl, 5 mM KCl, 2.5 mM CaCl₂, 100 μM MgCl₂). After adding 5 μl Annexin V-FITC, cells were incubated at room temperature for 15 min in the dark, propidium iodide (50 μg ml⁻¹) was added to each sample and cell death analysed by flow cytometry.

qPCR. Total RNA was extracted from cells with Trizol reagent (Invitrogen, 48190011), followed by isopropanol precipitation. We reverse-transcribed 2 µg total RNA to complementary DNA cDNA by using random hexamers and the M-MLV reverse transcriptase (Invitrogen, 28025-013) according to the manufacturer's protocol. qPCR with reverse transcription was performed using iTaq SYBR green supermix (Bio-Rad, 172-5852). The primers for different cytokines and chemokines used in this study are described in Supplementary Table 13.

Immunoblotting, immunoprecipitation and pull-down experiments. SDS-polyacrylamide gel electrophoresis and western blotting were performed in accordance with standard protocols. For signalling experiments, cells were lysed directly in Laemmli sample buffer. For co-immunoprecipitation experiments, HEK293T cells were co-transfected using Lipofectamine 2000. Cells were lysed in buffer B150 (20 mM Tris-HCl (pH 8.0), 150 mM KCl, 10% glycerol, 5 mM MgCl₂, and 0.1% NP-40) supplemented with a protease inhibitor tablet (Roche). Immunoassays were performed in RIPA buffer (10 mM Tris (pH 8.0), 150 mM NaCl, 1% Nonidet P-40, 0.1% SDS and 0.5% deoxycholate, supplemented with a protease inhibitor tablet (Roche)). In brief, whole cell lysates were collected 24 h later and incubated overnight with either M2 agarose beads (Sigma, A2220) or the immunoprecipitating antibody along with protein A/G sepharose beads (Sigma, P9424, P3296). Immunoprecipitates were eluted with Flag peptide (Sigma, F3290) or by boiling in Laemmli buffer. Eluted proteins were detected using monoclonal antibodies against green fluorescent protein, Flag, haemagglutinin or Myc tags. For direct pull-down assays, HEK293T cells were transfected with Myc-tagged NOD1, NOD2 or their deletion mutants. After 24 h, cells were lysed in WCE buffer (20 mM HEPES (pH 7.7), 500 mM NaCl, 1 mM EDTA, 0.25% Triton X-100, 1 mM EGTA) supplemented with 0.5 mM DTT and protease inhibitor cocktail (Roche). After extraction, protein samples were equilibrated in immunoprecipitation buffer (20 mM Tris (pH 8.0), 250 mM NaCl, 0.05% Nonidet P-40, 3 mM EDTA, 3 mM EGTA). Supernatants were collected and incubated for 16 h with purified recombinant histidine-tagged BID protein conjugated to Ni-NTA agarose beads (Qiagen). Immunoprecipitates were then analysed by western blotting. BID protein was produced in bacteria and purified by Ni-chelation chromatography similar to previous reports³⁶. For endogenous immunoprecipitation assays, 10⁸ BMDMs were used from wild-type, *Ripk2*^{-/-} or *Bid*^{-/-} mice. Cells were lysed in RIPA buffer supplemented with a protease inhibitor tablet (Roche), 5 mM sodium fluoride and 1 mM sodium orthovanadate. Lysates were suspended in 0.5 ml lysis solution, cleared by incubation with 15 µl protein A/G Sepharose beads, then incubated with 25 µl polyclonal antibody and 40 µl protein A/G beads at 4 °C overnight. Samples were then washed four times with lysis buffer, boiled in Laemmli buffer and analysed by SDS-polyacrylamide gel.

Intraperitoneal injection of LPS, MDP and γTri-DAP. Wild-type, RIP2-deficient and BID-deficient mice were injected intraperitoneally with 1 mg kg⁻¹ of γTri-DAP. Twenty-four hours after injection, infiltrating cells were collected in PBS by peritoneal lavage, counted and labelled for 30 min in the dark with FITC-labelled Ly6G/C antibody (BD Pharmingen, 553126). Gr-1⁺ cells were enumerated by flow cytometry using FACSCalibur (BD Biosciences). γTri-DAP (2 mg kg⁻¹) was injected intraperitoneally, blood was collected by heart puncture 2 h after agonist injection and serum used to quantify IL-6, KC and MCP-1 levels by ELISA. Similarly, MDP (100 µg; 500 µl per mouse) or LPS (10 mg kg⁻¹) were injected intraperitoneally. Blood was collected after 2 h and 4 h for MDP and 30 min for LPS. IL-6 levels were quantified by ELISA.

DSS colitis. Wild-type, *Ripk2*^{-/-} and *Bid*^{-/-} mice were given 4% DSS in drinking water for 6 days (days 0–5), then placed on regular water for 3 days (days 6–8). Mice were administered intraperitoneal MDP (100 µg in 500 µl PBS), poly I:C (100 µg in 500 µl PBS) or endotoxin-free PBS early during colitis on days 0, 1 and 2. Mice were monitored daily for body weight loss.

Histopathological analysis and immunohistochemistry. Colons were fixed in 10% formalin and embedded in paraffin. Serial sections (5 µm) were cut onto glass slides and stained with haematoxylin and eosin. Sections were blindly assessed for signs of colitis: tissue morphology, epithelial erosion, loss of goblet cells, loss of crypts, presence of inflammatory infiltrate and thickening of the colon wall.

Statistical analysis. A two-tailed Student's *t*-test was used to evaluate significant differences between two groups. One-way ANOVA was used to evaluate significant differences among multiple groups.

29. Ogura, Y. *et al.* Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-κB. *J. Biol. Chem.* **276**, 4812–4818 (2001).
30. Boutros, M., Bras, L. P. & Huber, W. Analysis of cell-based RNAi screens. *Genome Biol.* **7**, R66 (2006).
31. Zhang, J. H., Chung, T. D. & Oldenburg, K. R. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen.* **4**, 67–73 (1999).
32. Mi, H. *et al.* The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* **33**, D284–D288 (2005).
33. Rivals, I. *et al.* Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics* **23**, 401–407 (2007).
34. Eisen, M. B. *et al.* Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
35. Saldanha, A. J. Java Treeview-extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).
36. Zhai, D. *et al.* Humanin binds and nullifies Bid activity by blocking its activation of Bax and Bak. *J. Biol. Chem.* **280**, 15815–15824 (2005).

Discovery of novel intermediate forms redefines the fungal tree of life

Meredith D. M. Jones^{1,2}, Irene Forn³, Catarina Gadelha⁴, Martin J. Egan^{1,5}, David Bass², Ramon Massana³ & Thomas A. Richards^{1,2}

Fungi are the principal degraders of biomass in terrestrial ecosystems and establish important interactions with plants and animals^{1–3}. However, our current understanding of fungal evolutionary diversity is incomplete⁴ and is based upon species amenable to growth in culture¹. These culturable fungi are typically yeast or filamentous forms, bound by a rigid cell wall rich in chitin. Evolution of this body plan was thought critical for the success of the Fungi, enabling them to adapt to heterogeneous habitats and live by osmotrophy: extracellular digestion followed by nutrient uptake⁵. Here we investigate the ecology and cell biology of a previously undescribed and highly diverse form of eukaryotic life that branches with the Fungi, using environmental DNA analyses combined with fluorescent detection via DNA probes. This clade is present in numerous ecosystems including soil, freshwater and aquatic sediments. Phylogenetic analyses using multiple ribosomal RNA genes place this clade with *Rozella*, the putative primary branch of the fungal kingdom¹. Tyramide signal amplification coupled with group-specific fluorescence *in situ* hybridization reveals that the target cells are small eukaryotes of 3–5 µm in length, capable of forming a microtubule-based flagellum. Co-staining with cell wall markers demonstrates that representatives from the clade do not produce a chitin-rich cell wall during any of the life cycle stages observed and therefore do not conform to the standard fungal body plan⁵. We name this highly diverse clade the cryptomycota in anticipation of formal classification.

To investigate phylogenetic diversity among the deepest branches of the fungal tree of life, we aligned a broad selection of fungal small subunit ribosomal DNA (SSU rDNA) sequences using published phylogenies¹ as a guide, to which we added a comprehensive sampling of environmental DNA sequences available in GenBank^{6–8} (Supplementary Table 1). Our trees are consistent with previous analyses¹ demonstrating that nucleariids, a group of opisthokont amoebae⁹, branch as sisters to the fungi, indicating a transition from a phagotrophic to an osmotrophic form very early within the fungal radiation¹. Our analyses also recovered a highly diverse clade of environmental sequences branching with the fungi and demonstrated that current models of fungal evolution and biodiversity, which are largely based on cultured microbes, have missed a huge fraction of the kingdom (perhaps even approaching half). The analysis demonstrated preliminary bootstrap support (66/33%) for this large clade of environmental sequences branching with *Rozella*¹⁰, the putative primary branch in the fungal phylogeny¹. We name this clade cryptomycota (Fig. 1a).

The cryptomycota clade contains sequences recovered from diverse habitats and geographical locations, including soils, marine and freshwater sediments, freshwater planktonic samples and oxygen-depleted environments but it seems to be absent from samples of the upper marine water column (Fig. 1a, b). To investigate the ecology and cell biology of these deep-branching fungi, we designed several DNA probes for the detection of subgroups in the cryptomycota clade. Using the ARB program, with comprehensive GenBank and SILVA database sampling, we identified ten probes of about 18 base pairs (bp)

that are specific to different sequences in the cryptomycota clade; probes and their target sequences are listed in Supplementary Table 2. Two probes were used successfully as forward PCR primers in combination with a general eukaryotic SSU rDNA reverse primer, 1520r (ref. 8; see Supplementary Table 2 and Fig. 1c). We then used PCR to test for the presence of the cryptomycota sequences termed CM1 and CM2 in multiple samples from a local freshwater pond, three freshwater reservoirs (Dartmoor National Park) and four coastal marine surface water samples (Devon, UK). Of the primer sequences tested, CM1 and CM2 consistently amplified cryptomycota rDNA from the Washington Singer pond (Exeter University, Devon, UK, 50.7339 °N, 3.5375 °W). We constructed clone libraries from both sets of amplicons and sequenced 12 clones from each, recovering only sequences that were 99% similar to Washington Singer CM1 in the first library and to the Lily Stem CM2 sequence previously sampled from Priest Pot pond (Cumbria, UK, 54.372 °N, 2.990 °W) in the second. This process demonstrated that both probes, when used as forward PCR primers, are specific to the two target groups in the Washington Singer pond samples. We did not detect either subgroup in the marine waters tested; however, only 0.8% of the thousands of eukaryotic environmental sequences retrieved from oceanic surface waters are classified as belonging to the Fungi¹¹, indicating a low density of fungi cells in the upper marine water column.

We then aimed to increase gene sampling so that we could perform additional multi-gene phylogenetic analyses to test the branching position of the environmental sequences. We constructed environmental gene libraries to sample the wider SSU-ITS1–5.8S–ITS2–large subunit (LSU) gene array, using the group-specific forward PCR primers in combination with a general eukaryotic LSU reverse primer, 28Sr1 (ref. 12), see Fig. 1c and Supplementary Table 2). This allowed a second round of phylogenetic analysis with increased character sampling and resulted in improved topology support values throughout the phylogeny (Supplementary Fig. 2). Importantly, CM1 and CM2 sequences clustered as a sister clade to *Rozella* with 96% support using two distinct bootstrap methods (shown in red on Fig. 1a and in full on Supplementary Fig. 2). This confirms that the environmental sequences branch with the genus *Rozella*.

To visualize the cells of the novel clade, we used the ten cryptomycota probe sequences detailed in Supplementary Table 2 as HRP-conjugated probes for tyramide signal amplification fluorescence *in situ* hybridization (TSA-FISH). Of these probes, CM1.1 and CM1.2 were consistently successful when we applied a high-stringency method combined with multiple helper probes^{13,14}. We applied the TSA-FISH methods to the same water samples in which we had confirmed the presence of cryptomycota sequence using PCR (the Washington Singer pond and the Trenchford and Tottiford reservoirs sampled from Dartmoor National Park, UK). For both groups, our FISH analyses consistently recovered an ovoid cell type of 3–5 µm (Fig. 1d) among a dense 4', 6-diamidino-2-phenylindole (DAPI)-stained microbial community ($n = 98$, CM1.1; $n = 58$, CM1.2). In contrast, the other eight

¹School of Biosciences, University of Exeter, Exeter EX4 4QD, UK. ²Department of Zoology, Natural History Museum, Cromwell Road, London SW7 5BD, UK. ³Department of Marine Biology and Oceanography, Institut de Ciències del Mar, CSIC, Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain. ⁴Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP, UK. ⁵Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston Massachusetts 02115, USA.

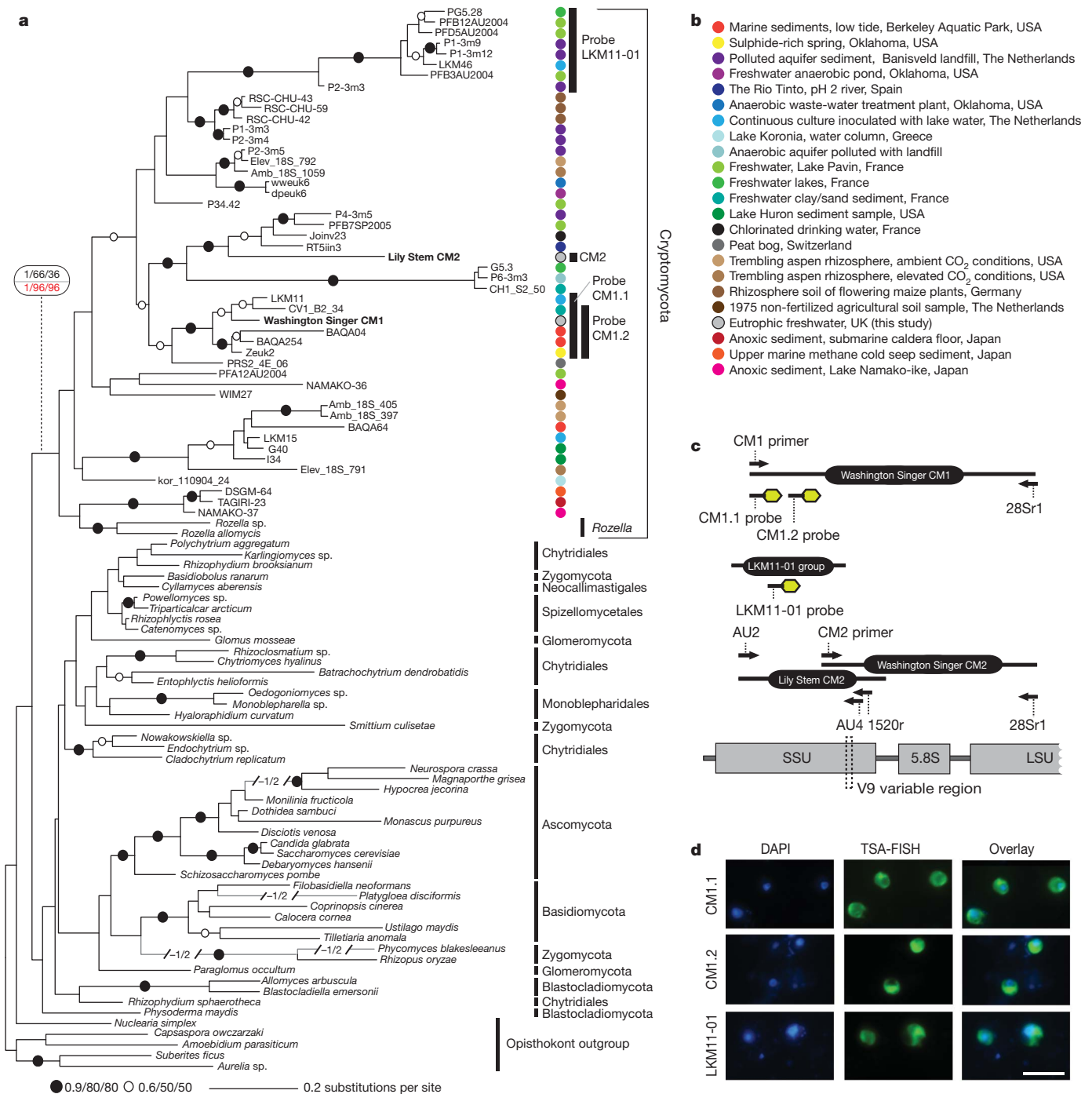


Figure 1 | Identification of the cryptomycota. **a**, Phylogeny demonstrates a diverse clade of environmental sequences (coloured dots) branching at the base of the Fungi. MrBayes tree topology was calculated from an alignment of 100 sequences and 1,012 DNA characters. Support values are summarized by black dots (indicating at least 0.9 Bayesian posterior probability and 80% bootstrap support by ML and Log-Det distance methods) or by black rings (Bayesian posterior probability above 0.6 and bootstrap support above 50%). The key node showing monophyly of cryptomycota encompassing *Rozella* is marked

cryptomycota group-specific probes and a panel of negative controls (detailed in the Methods) were consistently negative.

To complement this approach, we also used two LKM11 FISH probes previously used in association with additional eukaryotic probes to determine the eukaryotic community structure in freshwater environments¹⁵. LKM11 denotes a subsection of the environmental sequences branching within the cryptomycota clade detected in earlier studies^{6–8}. We used the probe LKM11-01 (Fig. 1a, c) to identify a

second subset of the cryptomycota clade and consistently recovered a cell type similar to that recovered for CM1 (Fig. 1d). The LKM11-02 probe was negative. We note that the abundance of cells identified by LKM11-01 was at least tenfold higher than observed for the CM1 cells, suggesting that the LKM11-01 subclade was more abundant than the CM1 subclade in the environments sampled.

To examine further the life cycle and morphology of the target groups in the freshwater samples, we tested whether the cells identified by

TSA-FISH possessed a flagellum. Flagella are a characteristic of chytrid fungi which, according to our phylogenetic analyses, branch closer to the cryptomycota than other fungal forms (Fig. 1a and Supplementary Figs 1 and 2). Co-localization of our TSA-FISH probes with the monoclonal antibody TAT1¹⁶ against α -tubulin (a major structural component of eukaryotic flagella) was used to test for flagellate forms in the cryptomycota. Evidence of zoospore (flagellate spore) formation was observed for both the CM1 and LKM11-01 subgroups (Fig. 2a and Supplementary Fig. 3). In a comparative FISH experiment using 12 hybridizations from four samples (two from the Washington Singer eutrophic pond and one each from the oligotrophic Trenchford and Tottiford reservoirs, all with three replicates and using the LKM11-01 probe) we found that 47–85% of cells possessed a single flagellum ($n = 455$ of a total of 696 cells). These data demonstrate that the cells are present in a zoospore form, actively growing, reproducing and seeking food. During a typical chytrid cell cycle, a flagellum is usually lacking during the cyst phase when a cell wall is formed^{5,17} and the cell enters dormancy; the majority of the remaining, unflagellate cryptomycota cells are therefore hypothesized to be cyst forms or cells in transition to form cysts (Fig. 2c).

We found preliminary evidence of a third phase of the cryptomycota life cycle. In both the LKM11-01 and CM1 groups, we observed non-flagellate cells attached to second-party cells (Fig. 2d, e) in what appeared to be an epibiotic (possibly parasitic or saprotrophic) association ($n = 39$ for LKM11-01, $n = 6$ for the CM1 probes). This represents a low total observation number ($n = 45$), an unavoidable

consequence of sampling interacting microbes directly from the environment. However, in support of the interaction hypothesis, the cryptomycota cells were frequently found on diatoms. Furthermore, we saw examples of diatom cells with several attached cryptomycota cells (Supplementary Fig. 4): this would be unlikely if cell attachment were a sampling artefact. In addition, we observed non-illuminated cells attached to diatoms, indicating that the probes target cryptomycota specifically and that this was not an auto-fluorescence artefact affecting attached cells generally. Because cryptomycota are currently uncultured, we have to rely on environmental observations alone to understand their biology; it is therefore likely that several aspects of their life cycle have not yet been observed. Consequently, key cellular apparatuses may remain undetected, such as germ tubes, filter-feeding structures, rhizoids, hyphae, sporangia and cell division characteristics. However, our TSA-FISH observations of three life stages enable us to propose a skeleton life cycle for the cryptomycota lineages sampled, although we note that because of the evolutionary diversity of cryptomycota, this model is unlikely to represent accurately the numerous unobserved forms that branch within the wider clade (Fig. 2f).

The evolution of a chitin-rich cell wall was one of the most important acquisitions which drove the success and diversification of the Fungi⁵, enabling these organisms to resist high osmotic pressure and feed by osmotrophy. *Rozella*, the only known genus branching within the cryptomycota clade (Fig. 1a), does not synthesize its own cell wall during many phases of its life cycle; instead, it appears to acquire cell

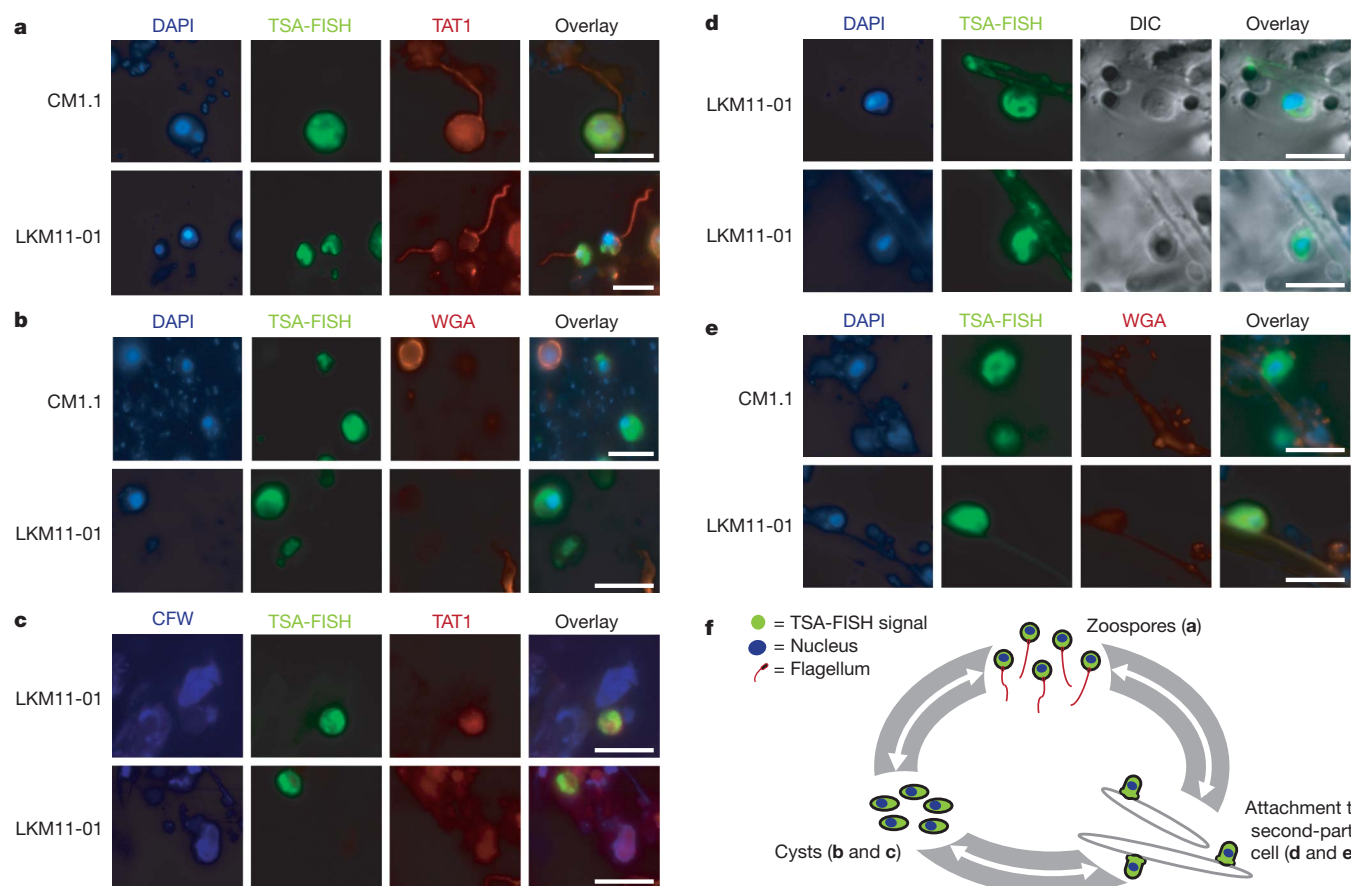


Figure 2 | Structural properties of cryptomycota cells and evidence for different life cycle stages. **a**, Micrographs showing flagella on cryptomycota cells, as detected by TAT1 tubulin antibody. **b**, Single cells without a chitin cell wall, as inferred by non-binding of wheat germ agglutinin (WGA). **c**, Non-flagellate putative cysts (lack of TAT1 signal indicates absence of flagellum), without a chitin/cellulose cell wall, as inferred by non-binding of calcofluor white. **d**, Cryptomycota cells attached to a second-party cell. Bright field differential interference contrast (DIC) shows filamentous structure of the second-party cell. **e**, Cryptomycota cells attached to a second-party cell; residual

staining identifies the boundary of the second-party cell and WGA identifies absence of chitin during attachment. Scale bars, 10 μ m. **f**, Putative cryptomycota skeleton life cycle (letters in brackets refer to micrographs a–e). This life cycle is limited to stages identified using TSA-FISH so additional stages are likely to remain unobserved (for example, sporangia stages and cell division) and the order of transition remains hypothetical. Furthermore, the diversity of the cryptomycota group strongly indicates that there are likely to be numerous life-cycle variations within the group, so this life cycle is unlikely to represent the wider diversity of cryptomycota.

wall material from its hosts during infection^{18,19}. To investigate the cell biology of the cryptomycota further, it was therefore important to determine whether they produce a chitin- and cellulose-rich cell wall like other fungi (Supplementary Fig. 5), or whether this character is missing, consistent with the biology of *Rozella*^{18,19}.

We combined a TSA-FISH-based strategy with fluorescent stains specific to key compounds of the fungal cell wall. Calcofluor white binds and illuminates chitin and cellulose, the rigidifying components of fungal cell walls⁵, and lectin wheat germ agglutinin binds and illuminates chitin²⁰. Using these stains, we readily observed non-cryptomycota cells surrounded by chitin and chitin/cellulose in our samples (Fig. 2b, c) and in additional control experiments (Supplementary Fig. 5), confirming that these methodologies readily detect fungal cell walls. However, we did not observe a chitin/cellulose cell wall in any of the three cryptomycota life-cycle stages identified by FISH (Fig. 2b, c and e; Supplementary Figs 6 and 7). The TSA-FISH-based strategy and our samples may miss some specific stages of the life cycle with a chitin/cellulose-rich cell wall, yet our analyses show that the cryptomycota form non-flagellate putative cysts and attach to other cells (with the flagellum absent), all without a detectable chitin/cellulose cell wall. These results indicate that the key characteristic defining the kingdom Fungi — a growth and development strategy supported by a rigid chitin-rich cell wall⁵ — is not present in key phases of the cryptomycota life cycle.

Using a combination of environmental DNA sequencing and fluorescence microscopy we have identified a new component of the fungal tree of life. We have tentatively named this wider group cryptomycota (crypto, hidden; mycota, of the kingdom Fungi) because a chitin-rich cell wall (one of the important fungal-defining characteristics) is so far undetected and because, with the exception of *Rozella*^{18,19}, the biology of this group is still largely cryptic. Significantly, the biodiversity within this clade is extensive, representing a breadth of rDNA molecular diversity similar to that of the currently sampled fungal kingdom.

METHODS SUMMARY

Phylogenetic analysis. The rDNA sequences from environmental samples and characterized species were obtained from GenBank, aligned using MUSCLE²¹ and then manually refined and masked using SEAVIEW²². Phylogenies were calculated using 1,000 fast-ML²³ bootstraps, 1,000 Log-Det²⁴ distance bootstraps and Bayesian²⁵ methods.

FISH identification of cryptomycota cells. Surface water samples (3–4 litres) were collected from freshwater and marine coastal sites around Devon (Supplementary Table 3) and prefiltered with Miracloth (pore size 22–25 µm, Merck Chemicals). Half of the volume was immediately filtered under vacuum through a 2.0-µm polycarbonate membrane filter (Millipore) for DNA extraction. The remaining half was fixed with formaldehyde (3.7% final volume) overnight before filtration and storage at –80 °C. Oligonucleotide probes were designed against several cryptomycota subclades (Fig. 1c and Supplementary Table 2), conjugated with the HRP enzyme and used for TSA-FISH according to ref. 14. FISH optimization was performed using a range of stringency conditions (see online Methods).

To investigate whether a flagellum was present, filter pieces subjected to TSA-FISH hybridization were re-permeabilized with 0.1% v/v nonidet P-40 in PBS (10 mM Na₂HPO₄, 2 mM KH₂PO₄, 137 mM NaCl, 2.7 mM KCl, pH 7.2), blocked with 1% w/v bovine serum albumin in PBS and incubated for 1 h with the monoclonal antibody TAT1¹⁶ against α-tubulin, followed by fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse immunoglobulins (Jackson ImmunoResearch/Strattech).

To determine the presence and composition of a cell wall, filter pieces pre-hybridized with TSA-FISH probes were counter-stained with cell wall markers: calcofluor white (Sigma, 1 mg ml^{–1}, fluorochrome for chitin and/or cellulose) and wheat germ agglutinin (Invitrogen, Alexa 555–580 tetramethylrhodamine conjugate, 5 µg ml^{–1}, for chitin). These were then viewed under epifluorescent ultraviolet radiation and red light respectively.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 July 2009; accepted 7 March 2011.

Published online 11 May 2011.

- James, T. Y. *et al.* Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**, 818–822 (2006).

- Pirozynski, K. A. & Malloch, D. W. The origin of land plants: a matter of mycotrophism. *Biosystems* **6**, 153–164 (1975).
- Wang, B. & Qiu, Y. L. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16**, 299–363 (2006).
- Hawksworth, D. L. The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycol. Res.* **105**, 1422–1432 (2001).
- Bartnicki-Garcia, S. *Evolutionary Biology of the Fungi* (eds Rayner, A. D. M., Brasier, C. M. & Moore, D.) 389–403 (Cambridge University Press, 1987).
- van Hatten, E. J., Mooij, W., van Agterveld, M. P., Gons, H. J. & Laanbroek, H. J. Detritus-dependent development of the microbial community in an experimental system: qualitative analysis by denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* **65**, 2478–2484 (1999).
- Lepère, C., Domaizon, I. & Debroas, D. Unexpected importance of potential parasites in the composition of freshwater small-eukaryote community. *Appl. Environ. Microbiol.* **74**, 2940–2949 (2008).
- Lefèvre, E. *et al.* Unveiling fungal zooflagellates as members of freshwater picoeukaryotes: evidence from a molecular diversity study in a deep meromictic lake. *Environ. Microbiol.* **9**, 61–71 (2007).
- Amaral Zettler, L. A., Nerad, T., O’Kelly, C. J. & Sogin, M. L. The nuclearioid amoebae: more protists at the animal-fungal boundary. *J. Eukaryot. Microbiol.* **48**, 293–297 (2001).
- Lara, E., Moreira, D. & Lopez-Garcia, P. Environmental clade LKM11 and *Rozella* form the deepest branching clade of Fungi. *Protist* **161**, 116–121 (2010).
- Massana, R. & Pedrós-Alió, C. Unveiling new microbial eukaryotes in the surface ocean. *Curr. Opin. Microbiol.* **11**, 213–218 (2008).
- Bass, D., Richards, T. A., Matthai, L., Marsh, V. & Cavalier-Smith, T. DNA evidence for global dispersal and probable endemicity of protozoa. *BMC Evol. Biol.* **7**, 162 (2007).
- Fuchs, B. M., Glöckner, F. O., Wulf, J. & Amann, R. Unlabeled helper oligonucleotides increase the in situ accessibility to 16S rRNA of fluorescently labeled oligonucleotide probes. *Appl. Environ. Microbiol.* **66**, 3603–3607 (2000).
- Kim, E. *et al.* Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. *Proc. Natl Acad. Sci. USA* **108**, 1496–1500 (2011).
- Mangot, J. F., Lepère, C., Bouvier, C., Debroas, D. & Domaizon, I. Community structure and dynamics of small eukaryotes targeted by new oligonucleotide probes: new insight into the lacustrine microbial food web. *Appl. Environ. Microbiol.* **75**, 6373–6381 (2009).
- Woods, A. *et al.* Definition of individual components within the cytoskeleton of *Trypanosoma brucei* by a library of monoclonal antibodies. *J. Cell Sci.* **93**, 491–500 (1989).
- Webster, J. & Weber, W. S. *Introduction to Fungi* 3rd edn (Cambridge University Press, 2007).
- Held, A. A. The zoospore of *Rozella allomycis*: ultrastructure. *Can. J. Bot.* **53**, 2212–2232 (1975).
- Held, A. A. *Rozella* and *Rozellopsis*: naked endoparasitic fungi which dress up as their hosts. *Bot. Rev.* **47**, 451–515 (1981).
- Bulawa, C. E. Genetics and molecular biology of chitin synthesis in fungi. *Annu. Rev. Microbiol.* **47**, 505–534 (1993).
- Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
- Galtier, N., Gouy, M. & Gautier, C. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.* **12**, 543–548 (1996).
- Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
- Lockhart, P. J., Steel, M. A., Hendy, M. D. & Penny, D. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11**, 605–612 (1994).
- Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank: N. J. Talbot for advice, K. Gull for the TAT1 antibody, L. Guillou for access to curated SSU database and the Broad Institute of the Massachusetts Institute of Technology and Harvard for making their *Rhizopus* and *Batrachomyces* genome sequence data publicly available. T.A.R. thanks the Leverhulme Trust for fellowship support. This work was primarily supported by a Natural Environment Research Council grant UK (NE/F011709/1). Additional support came from the Systematic Research Fund (awarded by the Systematics Association and the Linnean Society) to T.A.R., project FLAME (CGL2010-16304, MICINN) to R. M. and the BioMarkS project (European Funding Agencies from the ERA-net program BiodivERsA) to T.A.R. and R.M.

Author Contributions This study was conceived by T.A.R. and M.D.M.J. with assistance from D.B. and R.M. M.D.M.J. performed the molecular biology experiments with assistance from I.F. (FISH), C.G. (immunolocalization) and M.J.E. (microscopy). T.A.R. performed the bioinformatics and phylogenetic analysis. T.A.R. and M.D.M.J. wrote the paper with assistance from D.B. and R.M.

Author Information Novel sequence data have been deposited in GenBank under accession numbers FJ687265, FJ687267 and FJ687268. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.A.R. (thomr@nhm.ac.uk).

METHODS

DNA extraction, PCR and environmental gene library construction. DNA was extracted from an epiphytic microbial community scraped from a submerged lily stem (Priest Pot pond, 54.372°N, 2.990°W) using the maximum yield protocol of the MoBio UltraClean Soil DNA Extraction Kit (Cambio), after filtration of a 50-ml stem-scrapping suspended in Priest Pot water onto a GF-F filter (Whatman). Amplified PCR products of lily stem-scrappings (using the AU2 and AU4^{26,27} primers, see Supplementary Table 2) were excised from the gel, purified (Wizard SV Gel and PCR Clean-Up System, Promega) and cloned (StrataClone, Stratagene) as suggested by the manufacturer. Plasmid purification (Wizard plus Miniprep Kit, Promega) was performed on 24 colonies before sequencing in the forward direction (Cogenics). Chromatograms were checked manually using Sequencher (Genecodes) and used for BLASTn searches of the National Center for Biotechnology Information, after which one positive cryptomycota clone was chosen for double-strand sequencing (GenBank number FJ687267).

Ten novel cryptomycota probes/primers (Supplementary Table 2), designed using the ARB²⁸ program (using database sampling from Silva²⁹ and GenBank³⁰), were tested for specificity on community DNA extracted from the sites detailed in Supplementary Table 3 by using PCR with a forward-priming version of the probe in combination with the universal eukaryotic SSU reverse primer 1520r⁸ (Supplementary Table 2). Two probes/primers were successful, yielding PCR products of the correct estimated amplicon length from Washington Singer environmental DNA using PCR amplification protocol 1 (Supplementary Table 4). The CM1–1520r PCR amplicon size was ~1,600 bp and the CM2–1520r PCR amplicon size was ~420 bp. Both amplicons crossed the variable V9 region (Fig. 1c). Two clone libraries were constructed and 12 clones from each library were sequenced.

Targeting longer rDNA sequences. We used the CM1 forward primer and an extended version of the CM2 forward primer in combination with the reverse LSU primer 28S1¹² (primer sequences given in Supplementary Table 2) to enable rRNA gene array PCR amplification. The CM1 and CM2 SSU–ITS1–5.8S–ITS2–LSU PCR amplifications were conducted using PCR protocols 2 and 3 respectively (Supplementary Table 4). Amplicons were purified, cloned as before and representatives of the CM1 (2,600 bp) and CM2 (1,449 bp) groups were double-strand sequenced. The Washington Singer CM2 group SSU-to-LSU sequence overlapped the Lily Stem CM2 SSU sequence by 316 identical bp, which encompassed the variable V9 region (Fig. 1c). For phylogenetic analyses, these sequences were concatenated, generating a sequence of 2,711 bp.

Sequence analysis, alignment and phylogenetic analyses. An SSU multiple-sequence DNA alignment was constructed using the work of James *et al.*^{1,31} as a guide, with taxon sampling focusing on ‘deep’ branching fungal groups and on taxa with five rDNA sequence regions (SSU, ITS1, 5.8S, ITS2 and LSU) available (Supplementary Table 5). For the first alignment, we used the Washington Singer CM1 and Lily Stem CM2 SSU sequences as seeds for BLASTn searches of the GenBank non-redundant database to identify additional SSU sequences not sampled among currently published fungal phylogenies (for example, refs 1, 31). The alignment was calculated using the automatic alignment program MUSCLE²¹, then manually refined and masked using SEAVIEW²². To check that we had included all available SSU sequences, we searched the GenBank non-redundant database again using all the members of the cryptomycota clade and the *Rozella* sequences as BLASTn search seeds (searches conducted May 2010). Many of the environmental sequences recovered were too short to be included in our phylogenetic analyses (Supplementary Table 1) but they indicated that the diversity of the cryptomycota clade is very extensive and even more ecologically varied than shown in Fig. 1a, b. The final phylogenetic analysis was based on environmental SSU sequences that overlapped both the CM1 and CM2 probe-binding regions; that is, only environmental SSU rDNA sequences of more than ~1,350 bp (BLAST searches conducted in May 2010). We removed all sequences with evidence of chimaeras identified by comparison to the curated SSU rDNA database (gift of L. Guillou).

A second multiple-sequence alignment was generated on the basis of the SSU–ITS1–5.8S–ITS2–LSU DNA sampling often used in the fungal tree of life project^{1,31} and including our two cryptomycota sequences. This was processed, as described above, to generate a 54-sequence and 1,877-character data matrix with the ITS1 and ITS2 regions masked out because they were too variable for use. The alignment was checked for patterns of sequence variation that could be the products of chimaeras, both by visual analysis of signature sequences across the alignment and by constructing trees from one-sixth divisions of the masked alignment.

Both the SSU and SSU–5.8S–LSU sequence alignments were run through the program MODELGENERATOR³² to identify the most appropriate model parameters for phylogenetic analysis (SSU: model of substitution GTR³³ + Γ (0.28, eight rate categories); SSU–5.8S–LSU: model of substitution GTR³³ + Γ (0.35, eight rate categories + I = 0.22)). Bayesian tree topologies were inferred for both DNA alignments using a Metropolis-coupled Markov chain Monte Carlo (MCMCMC) method. For this, we used MrBayes²⁵ with two runs, each with four

Markov chains for 5,000,000 generations with default ‘temperature’ settings, a sampling frequency of 250 generations, six substitution categories (nst = 6) and an eight-category gamma model with spatial autocorrelation between rates at adjacent sites (rates = adgamma) and a covarion-like model (covarion = yes). The MCMCMC log likelihood results were compared and a ‘burn-in’ of 1,000 generations (SSU–5.8S–LSU analysis) and 2,500 generations (SSU analysis) was removed to sample only trees from the stationary phase of the MCMCMC searches. The resulting samples of trees were then used to construct the majority-rule consensus trees. Tree topologies were evaluated using two bootstrap methods. For each phylogeny, 1,000 bootstraps were calculated using the fast-ML method implemented in the PHYLML program²³ using the model parameters from the MODELGENERATOR analyses. As compositional bias can cause phylogenetic reconstruction artefacts³⁴, we ran an additional bootstrap analysis using Log-Det methods²⁴ with stepwise addition (ten random starting trees per replicate) and a tree-bisection-reconnection branch-swapping algorithm. In both Log-Det bootstrap analyses, the ‘proportion of invariant sites’ parameter was estimated using PAUP³⁵: SSU = 0.457 and SSU–5.8S–LSU = 0.509.

FISH identification of cryptomycota cells. Surface water samples (3–4 litres) were collected from freshwater and marine coastal sites around Devon (Supplementary Table 3) and prefiltered (pore size 22–25 μ m) with Miracloth (Merck Chemicals). Half of the volume was immediately filtered under vacuum through a 2.0- μ m polycarbonate membrane filter (Millipore) for DNA extraction. The remaining half was fixed with formaldehyde (3.7% final volume) overnight before filtration and storage at –80 °C. Oligonucleotide probes were designed against several cryptomycota subclades (Fig. 1a, c and Supplementary Table 2), conjugated with the HRP enzyme and used for TSA-FISH^{13,14}. FISH optimization was performed using a range of stringency conditions (see below).

FISH stringency, helper probe, chitin wash and control experiments. The thirteen FISH probes, including a negative control probe (Supplementary Table 2), conjugated with the HRP enzyme, were used for TSA-FISH according to the method detailed in ref. 14. Initial hybridizations were at 35 °C with 30% formamide hybridization buffer. These were unsuccessful in revealing our target group. Formamide concentrations (30%, 40%, 45% and 50%) were varied in combination with a range of hybridization temperatures (35, 42 and 46 °C). Probes CM1.1 and CM1.2, both with a minimum of two mismatches to all non-target database sequences, were successful in illuminating candidate cryptomycota cells. These cells were not seen in the negative controls (detailed below).

To improve the rate of detection further, we designed additional helper probes (without the HRP conjugate) specific to the CM1.1 and CM1.2 target groups, which bind either side of the TSA-FISH probe-binding site (Supplementary Table 2). This approach should theoretically open up the tertiary structure of the ribosome and increase accessibility for the TSA-FISH probe to the target binding site^{13,14}. Group-specific helper probes (Supplementary Table 2) were used in conjunction with the addition of 10% dextran sulphate to the hybridization reaction. For both probes, the greatest number of target cells was viewed using a 30% formamide hybridization buffer at 42 °C, with the addition of 10% dextran sulphate and helper probes.

We then combined our TSA-FISH protocol with a chitinase incubation step before hybridization, to test whether a chitin cell wall was inhibiting detection of the target cell³⁶. Filter pieces were incubated in a 1 mg ml^{–1} solution of chitinase (Sigma) in modified PBS (10 mM Na₂HPO₄, 2 mM KH₂PO₄, 137 mM NaCl, 2.7 mM KCl, pH 5.5 rather than 7.2) with 1% SDS. Filters were incubated for 10 min at 30 °C, rinsed with distilled water and subjected to the TSA-FISH protocol as already described. The chitinase wash step did not appear to increase detection frequency and therefore a chitin cell wall was not considered an inhibitor of detection. The chitinase wash step was not used during further TSA-FISH analyses.

For the LKM11-01 probe, we used the methodology described in ref. 15 with filter pieces hybridized at a range of formamide concentrations (30%, 40%, 45% and 50%). The optimum concentration was 40% and this was used for all subsequent hybridizations. Again, for LKM11-01 TSA-FISH, six replicate preparations were completed using the chitinase wash step. These extra analyses again failed to increase detection frequency, indeed it often reduced the frequency of detection in replicate filters indicating that inhibition of TSA-FISH binding by a chitin cell wall was not a factor, and again this protocol step was dropped from further experiments.

Negative control tests were devised to check the specificity of the probes and eliminate the possibility of autofluorescence in the sample. Filter pieces from each sample were hybridized in identical conditions to those described above, but without a TSA-FISH probe, which was replaced with distilled water. Furthermore, hybridizations conducted using a TSA-FISH probe specific to a 16S plastid SSU rRNA gene (see Supplementary Table 2) for a novel algal group¹⁴ and probes for further cryptomycota groups were used as additional negative controls. Positive cells were never observed.

Co-staining for flagella and cell wall apparatuses. To investigate whether a flagellum was present, filter pieces subjected to TSA-FISH hybridization were re-permeabilized with 0.1% v/v nonidet P-40 in PBS (10 mM Na₂HPO₄, 2 mM KH₂PO₄, 137 mM NaCl, 2.7 mM KCl, pH 7.2), blocked with 1% w/v bovine serum albumin in PBS and incubated for 1 h with the monoclonal antibody TAT1¹⁶ against α -tubulin, followed by fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse immunoglobulins (Jackson ImmunoResearch/Strattech).

To determine the presence and composition of a cell wall, filter pieces pre-hybridized with TSA-FISH probes were counter-stained with cell wall markers: calcofluor white (Sigma, 1 mg ml⁻¹, fluorochrome for chitin and/or cellulose) and wheat germ agglutinin (Invitrogen, Alexa 555–580 tetramethylrhodamine conjugate, 5 μ g ml⁻¹, for chitin). These were then viewed under epifluorescent ultraviolet radiation and red light respectively. These markers readily detected cell wall structures in a *Blastocladiella emersonii* culture (Supplementary Fig. 5) and among the total microbial population sampled on our environmental filters (Fig. 2b, c and Supplementary Figs 6 and 7).

26. Bass, D. *et al.* Yeast forms dominate fungal diversity in the deep oceans. *Proc. R. Soc. Lond. B* **274**, 3069–3077 (2007).
27. Vandenkoornhuyse, P., Baldauf, S. L., Leyval, C., Straczek, J. & Young, J. P. W. Extensive fungal diversity in plant roots. *Science* **295**, 2051 (2002).
28. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
29. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).
30. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. GenBank. *Nucleic Acids Res.* **34**, D16–20 (2006).
31. James, T. Y. *et al.* A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia* **98**, 860–871 (2006).
32. Keane, T. M. *et al.* Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* **6**, 29 (2004).
33. Lanave, C., Preparata, G., Saccone, C. & Serio, G. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* **20**, 86–93 (1984).
34. Foster, P. G. & Hickey, D. A. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J. Mol. Evol.* **48**, 284–290 (1999).
35. Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods)*. Version 4 (Sinauer Associates, 2002).
36. Baschien, C., Manz, W., Neu, T. R. & Szewzyk, U. Fluorescence in situ hybridization of freshwater fungi. *Int. Rev. Hydrobiol.* **86**, 371–381 (2001).

Glutamate induces *de novo* growth of functional spines in developing cortex

Hyung-Bae Kwon & Bernardo L. Sabatini

Mature cortical pyramidal neurons receive excitatory inputs onto small protrusions emanating from their dendrites called spines. Spines undergo activity-dependent remodelling, stabilization and pruning during development, and similar structural changes can be triggered by learning and changes in sensory experiences^{1–4}. However, the biochemical triggers and mechanisms of *de novo* spine formation in the developing brain and the functional significance of new spines to neuronal connectivity are largely unknown. Here we develop an approach to induce and monitor *de novo* spine formation in real time using combined two-photon laser-scanning microscopy and two-photon laser uncaging of glutamate. Our data demonstrate that, in mouse cortical layer 2/3 pyramidal neurons, glutamate is sufficient to trigger *de novo* spine growth from the dendrite shaft in a location-specific manner. We find that glutamate-induced spinogenesis requires opening of NMDARs (*N*-methyl-D-aspartate-type glutamate receptors) and activation of protein kinase A (PKA) but is independent of calcium-calmodulin-dependent kinase II (CaMKII) and tyrosine kinase receptor B (TrkB) receptors. Furthermore, newly formed spines express glutamate receptors and are rapidly functional such that they transduce presynaptic activity into postsynaptic signals. Together, our data demonstrate that early neural connectivity is shaped by activity in a spatially precise manner and that nascent dendrite spines are rapidly functionally incorporated into cortical circuits.

During postnatal development, the formation and elimination of glutamatergic synapses are thought to be reflected in the growth and retraction of dendritic spines. In cortical pyramidal neurons, waves of new spine growth (spinogenesis) and synapse formation (synaptogenesis) occur at specific developmental stages, followed by pruning as the brain matures⁵. Many signals have been proposed to trigger and regulate *de novo* spine growth in a developing circuit including neurotrophins, neurotransmitters and cell-adhesion molecules^{6–9}. To uncover the triggers for and mechanisms of spinogenesis, we imaged dendrites of enhanced green fluorescent protein (EGFP)-expressing cortical layer 2/3 pyramidal neurons while releasing glutamate at a specific dendritic location by two-photon laser-induced photolysis of (4-methoxy-7-nitroindolyl)-glutamate (MNI-glutamate) (Fig. 1). Analysis was performed in acute cortical brain slices from young mice (postnatal day (P) 8–12), a period in which spinogenesis occurs *in vivo*¹⁰.

Stimulation near the edge of a dendrite with 40 0.5-ms laser pulses at 0.5 Hz in a Mg^{2+} -free extracellular solution induced growth of a new spine in approximately 14% of cases (Fig. 1a–d and Supplementary Fig. 1), showing the possibility of *de novo* spinogenesis induced by glutamate exposure¹¹. Increasing stimulation frequency and laser pulse duration while maintaining the total number of stimuli at 40 increased the rate of spinogenesis such that, at 5 Hz with 4 ms duration, a maximal success rate of approximately 50% was achieved (Fig. 1c). Nascent spines arose from the dendrite where glutamate was released with high specificity (Fig. 1b) such that more than 70% of them grew within 1 μm of the uncaging spot (Fig. 1e) and 94% of them grew on the side of the dendrite exposed to glutamate.

In 128 of 132 examples of glutamate-induced spinogenesis, the spine was seen to emerge without a filopodial stage (see Supplementary Fig. 2a for an exception). Instead, spine growth occurred incrementally but explosively such that the spine head volume increased from 10 to 90% of maximum within 11.8 ± 1.5 pulses of glutamate (5.9 ± 0.8 s at 2 Hz stimulation) (Fig. 2a–c and Supplementary Fig. 3). The final sizes and lengths of the newborn spines were heterogeneous but not different from those of pre-existing neighbouring spines (Fig. 2d, e). The lifetime of newly formed spines was variable such that approximately 20% lasted less than 2 min but those that lasted 5 min were stable and remained for at least 30 min (Supplementary Fig. 4). Thus, these newly formed spines either did not require continued exposure to glutamate for maintenance or they received glutamate from an alternative source such as an axonal bouton.

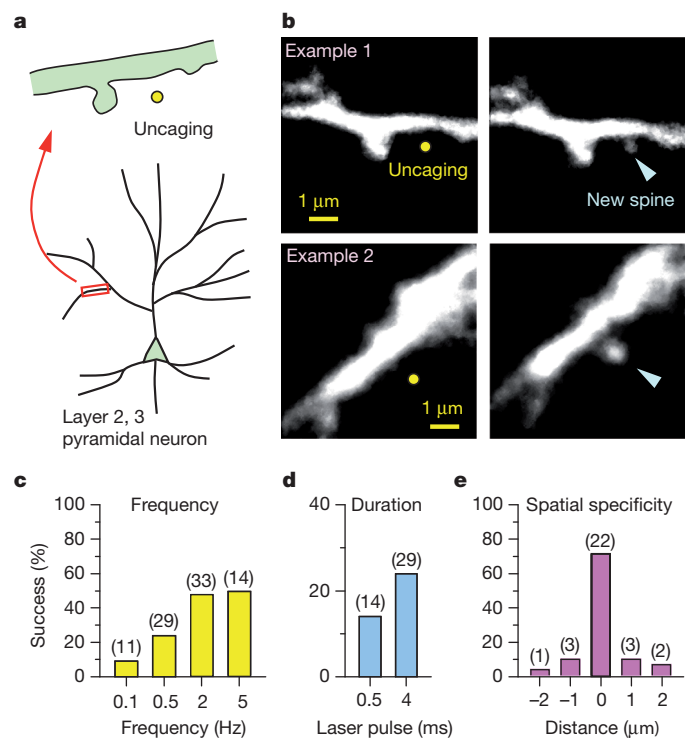


Figure 1 | *De novo* spine generation is induced by glutamate uncaging.

a, Dendrites of EGFP-expressing neurons in acute slices from P8–12 mice were visualized with two-photon laser scanning microscopy, and glutamate was released by photolysis of caged glutamate near a low-spine density section of dendrite. b, Examples of *de novo* spine formation induced by photolytic release of glutamate (40 pulses of MNI-glutamate uncaging at 2 Hz in Mg^{2+} -free artificial cerebrospinal fluid). Yellow circles, the uncaging spots; arrowheads, new spines. c–e, Most new spines grew near the uncaging spot and the success percentage depended on the frequency (c, laser pulse duration = 4 ms) and duration (d, stimulation frequency = 0.5 Hz) of glutamate uncaging. Experiment numbers for each bar are indicated in parentheses.

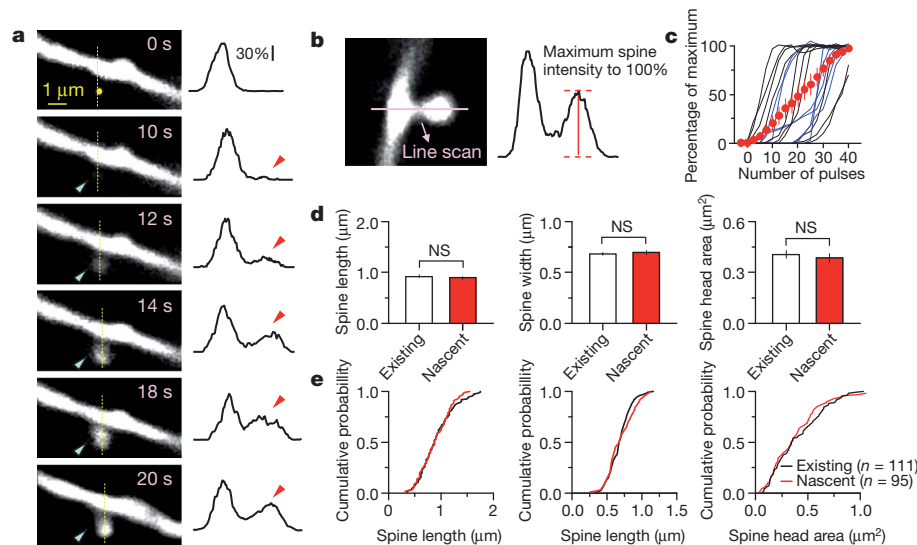


Figure 2 | New spines grow rapidly and acquire morphology similar to pre-existing spines. **a**, Left, time-lapse images of spine formation during glutamate uncaging (40 pulses, 2 Hz) showing the uncaging spot (yellow circle) and nascent spine (blue arrowhead). Right, fluorescence intensity profiles along the yellow line reveal that the spine head fluorescence increases gradually but rapidly (red arrowhead). **b**, Illustration of the measurement of spine head fluorescence during spinogenesis as a percentage of the maximum fluorescence intensity reached. **c**, Time course of individual (black, 2 Hz; blue, 0.5 Hz) and

average (red) fluorescence intensity increases during spinogenesis ($n = 17$). Error bars, s.e.m. **d**, Average of apparent spine length, width and head area from nascent ($n = 95$) and neighbouring existing ($n = 111$) spines (existing and nascent: length: $0.92 \pm 0.03 \mu\text{m}$, $0.89 \pm 0.03 \mu\text{m}$, $P > 0.1$; width: $0.68 \pm 0.02 \mu\text{m}$, $0.70 \pm 0.02 \mu\text{m}$, $P > 0.1$; head area: $0.41 \pm 0.02 \mu\text{m}^2$, $0.38 \pm 0.03 \mu\text{m}^2$, $P > 0.1$). **e**, Cumulative distributions demonstrating that the morphologies of pre-existing and nascent spines are not different.

Glutamate-induced spinogenesis was restricted within postnatal developmental such that its efficiency diminished by P14–15 and it failed to occur by P19–20 (Supplementary Fig. 5). This was not due to decreased glutamate receptor activation in older animals because the uncaging-evoked excitatory postsynaptic current (uEPSC) was larger at P19–20 than at P10–12 (Supplementary Fig. 6).

Previous ultrastructural studies have revealed a high frequency of dendrite shaft synapses in hippocampus in early postnatal life that decreases as spinogenesis occurs¹², leading to a model of synaptic development in which synapses are initially formed directly onto the dendritic shaft and a spine subsequently grows from this point with the synapse attached. On the other hand, rapid movement of a physically connected spine head and axonal bouton together through a complex neuropile is difficult to reconcile with the high density of crossing axons and dendrites¹³. Our data demonstrate that glutamate uncaging-induced spinogenesis occurs with high spatial specificity and probability on the side of the dendrite exposed to glutamate. These findings place a lower limit of approximately $1 \mu\text{m}^{-1}$ for the density of dendritic shaft synapses required to support this model of spinogenesis. To estimate the number of dendritic shaft synapses, Ca^{2+} imaging was performed in conditions in which most synapses formed onto a stretch of dendrite were activated (approximately 90%, Supplementary Fig. 7). Under these conditions, we observed hotspots of Ca^{2+} influx in spineless stretches of dendrite at a density of $0.05 \mu\text{m}^{-1}$. This corresponds to a density of dendritic shaft synapses containing NMDARs that is approximately 20-fold less than necessary to explain the specificity and efficiency of glutamate-induced spinogenesis (see Supplementary Fig. 7 for further discussion).

The high success rate of spinogenesis induced by glutamate uncaging allows identification of the signalling pathways that couple activity to spine growth (Fig. 3). Previous analyses of spine generation induced by electrical stimulation indicate a requirement for NMDARs in this process^{14,15}, and at this age NMDARs are found throughout the dendrite (Supplementary Fig. 8). Preventing NMDAR activation with the antagonist 3-(2-carboxypiperazin-4-yl)propyl-1-phosphonic acid (CPP) nearly abolished spinogenesis whereas it was unaffected by inhibiting α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA)/

kainate glutamate receptors with 2,3-dihydroxy-6-nitro-7-sulfamoyl-benzo[f]quinoxaline-2,3-dione (NBQX) (Fig. 3a). The voltage-gated sodium channel antagonist tetrodotoxin also did not affect spinogenesis, discarding the possibility that postsynaptic action potentials are necessary. Addition of extracellular Mg^{2+} significantly decreased the success rate, suggesting that the degree of current flux through NMDARs plays a crucial role in triggering spine formation. Blocking either mGluR1 or mGluR5 using 7-hydroxyiminocyclopropan[b]chromen-1a-carboxylic acid ethyl ester (CPCCOEt) or 2-methyl-6-(phenylethynyl)-pyridine (MPEP) or with the less selective group I mGluR antagonist 1-aminoindan-1,5-dicarboxylic acid (AIDA) demonstrated that neither was strictly necessary for spinogenesis. Lastly, depleting Ca^{2+} stores with cyclopiazonic acid or thapsigargin significantly inhibited spine formation. Thus our data indicate that NMDARs, with additional contributions from intracellular stores provide, the coupling between glutamate and activation of intracellular pathways responsible for spinogenesis.

We further considered intracellular signalling pathways that might be activated in a spatially delimited fashion by glutamate and could provide the spatial information necessary for local spine growth. Previous studies of long-term potentiation and associated spine enlargement in older animals demonstrated a need for activation of CaMKII or signalling by neurotrophin receptor tyrosine kinases such as the BDNF receptor TrkB^{16–19}. However, we found that activity-dependent spinogenesis is unaffected by the kinase inhibitors KN-62, KN-93 and K252a, indicating independence from CaMKII and TrkB signalling (Fig. 3a).

The cyclic AMP (cAMP)-activated kinase PKA is required for long-term potentiation induction in younger neurons²⁰, and gradients in its concentration can be maintained over micrometre-length scales²¹. Raising cAMP concentration by applying the adenylate cyclase activator forskolin was not sufficient to generate spines on its own ($n = 3$, data not shown), but glutamate uncaging in its presence increased the spinogenesis success rate to approximately 80% (Fig. 3a). In the presence of forskolin, multiple spines often grow with each induction attempt (Fig. 3b, c), including at sites more distant from the uncaging location (Fig. 3d). In addition, the PKA inhibitor H-89 prevented new

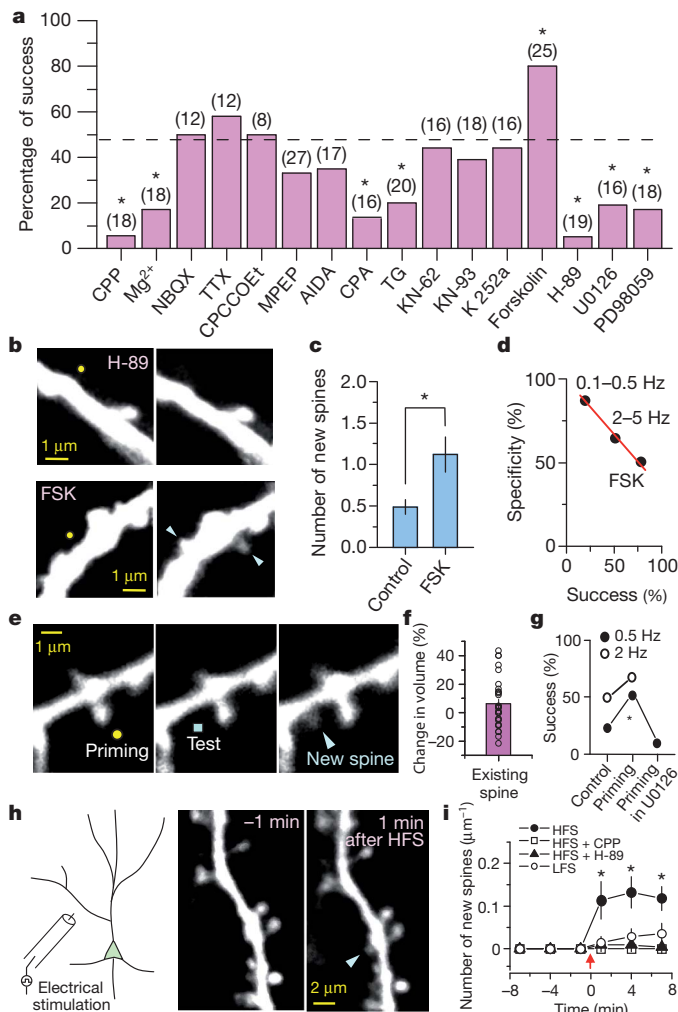


Figure 3 | Molecular mechanisms of glutamate-induced spine formation. **a**, Spine formation using the 40 pulses at 2 Hz protocol was tested in the presence of pharmacological agents. The dotted line indicates the success percentage in control conditions with which statistical comparison was made. The number of induction attempts for each condition is given in parentheses, and the numbers of successes and total trials are summarized in Supplementary Table 2. TG, thapsigargin; CPA, cyclopiazonic acid. **b**, Examples of blockade of spine generation and of exuberant spine growth. Images taken before (left) and after (right) glutamate uncaging in the presence of either H-89 or forskolin (FSK). Arrowheads, nascent spines. **c**, Summary of the average number of new spines with each induction attempt (control: 0.48 ± 0.09 , $n = 33$; FSK: 1.12 ± 0.21 , $n = 25$, $P < 0.005$). **d**, Inverse relation between the location specificity and success rate. Data from three groups (0.1–0.5 Hz, 2–5 Hz and FSK at 2 Hz) are plotted. Specificity was measured as the percentage of cases in which the spine arose within 1 μm of the uncaging spot. **e**, Representative images of priming experiments in which 40 pulses of 2 Hz glutamate uncaging were delivered to a pre-existing spine (yellow circle) followed by an additional 40 pulses (0.5 Hz or 2 Hz) delivered to the nearby dendrite (blue square). Releasing glutamate did not cause enlargement of the pre-existing spine head (left), but did trigger new spine growth from the dendrite (middle, right). **f**, Changes in the fluorescence of pre-existing spine heads exposed to the priming stimulus (individual spines: circles; bar graph: average ± s.e.m.) ($n = 31$). **g**, Percentage of successful spine generation at the indicated test frequencies with and without priming. U0126 prevented spinogenesis facilitated by the priming protocol. **h**, The experiment (left) and images of a dendrite 1 min before (middle) and after (right) HFS (2×100 pulses at 100 Hz). **i**, Average numbers of new spines generated by HFS per micrometre of dendrite in control conditions (at 7 min: 0.12 ± 0.03 , $n = 11$, $P < 0.005$), in the presence of CPP (0, $n = 8$, $P > 0.5$), or H-89 (0.004 ± 0.001 , $n = 12$; $P > 0.5$). The same number of pulses at a lower frequency (LFS, 10 Hz) generated fewer new spines (0.035 ± 0.024 , $n = 7$, $P > 0.1$). Error bars, s.e.m.

spine growth (Fig. 3a), indicating that PKA activity is necessary but not sufficient for spinogenesis.

The small guanosine triphosphatase Ras is activated by Ca²⁺ influx through NMDARs and signals through mitogen-activated protein kinase (MAPK) to promote long-term potentiation^{17,22}. We found that the MAPK pathway was necessary for spinogenesis because the success rate was significantly reduced by blocking the upstream activator MAPK kinase 1/2 (MEK1/2) with U0126 or MEK1 with PD98059 (Fig. 3a). In pyramidal neurons, activated Ras diffuses from active spines to neighbouring spines and heterosynaptically facilitates plasticity¹⁷. To examine if a similar phenomenon potentiates spinogenesis, we delivered a 'priming' stimulus to a pre-existing spine (40 pulses at 2 Hz, Fig. 3e) and then a 'test' stimulus to the dendritic shaft within 1–2 min. In contrast to previous studies of older neurons in hippocampus^{16–18}, we found no consistent increase in volume of the existing spine in response to this priming stimulus (Fig. 3f), supporting the idea that spinogenesis is not simply due to the growth of an undetectable pre-existing spine. Nevertheless, the success rate of spine generation was enhanced by the priming stimulus in a MEK1/2-dependent manner such that the low level of spinogenesis evoked by a 0.5 Hz test stimulus was increased. In contrast, the priming stimulus did not increase the success rate when the test stimulus was delivered at 2 Hz (Fig. 3g), indicating that the priming protocol shifted the induction threshold for spinogenesis.

With the exception of the NMDAR and PKA antagonists, none of the pharmacological manipulations that altered spinogenesis rates affected dendritic currents or Ca²⁺ transients (Supplementary Fig. 8). As expected, CPP largely abolished the Ca²⁺ transient and the prolonged phase of the currents. Consistent with a facilitation of Ca²⁺ influx through NMDARs by PKA^{23,24}, H-89 lowered dendritic Ca²⁺ transients by approximately 20% (Supplementary Table 1). However, this effect is insufficient to explain the near abolition of spinogenesis because the rate of spinogenesis evoked by 0.5 ms uncaging pulses in control conditions was higher than that evoked by 4 ms pulses in H-89, despite eliciting smaller dendritic Ca²⁺ transients (Supplementary Fig. 8 and Supplementary Table 1).

We examined whether endogenous synaptic activity generates new spines in cortical tissue from young mice through similar pathways. In normal extracellular Mg²⁺, high-frequency (100 pulses at 100 Hz, delivered twice, separated by 10 s) but not low-frequency (10 Hz) electrical tetani rapidly triggered new spine growth (Fig. 3h, i). Blockade of NMDARs or PKA prevented this synaptically evoked spine growth (Fig. 3i) as well as spontaneous new spine growth that occurred when NMDAR activation was increased by removing extracellular Mg²⁺ and blocking inhibitory GABA_A receptors (Supplementary Fig. 9). Hence, multiple experimental models demonstrate that activity-dependent spinogenesis in developing cortex requires NMDAR- and PKA-dependent signalling.

To determine if nascent spines detect synaptically released glutamate and are functionally incorporated into the circuit, we generated new spines by high-frequency stimulation (HFS) and examined their synaptic responses using optical quantal analysis of synaptic properties. A whole-cell recording was obtained and the probability and amplitude of synaptically evoked NMDAR-mediated Ca²⁺ transients in the spine head were monitored (Fig. 4a). Using the Ca²⁺-sensitive green fluorophore Fluo-5F, we detected stimulus-evoked Ca²⁺ transients in the heads of newly grown spines (Fig. 4b, c), demonstrating that they sense synaptic activity within a neural circuit within 30 min after growth. Similar results were obtained in five of seven new spines and in 11 of 16 pre-existing spines. Analysis of the spines in which an evoked Ca²⁺ transient could be detected indicated that nascent spines displayed smaller and less frequent synaptically evoked Ca²⁺ transients than pre-existing spines (Fig. 4d, e). Similar results were obtained for new spines that grew in response to glutamate uncaging and were probed using a glass-stimulating electrode placed near the spine (Supplementary Fig. 10).

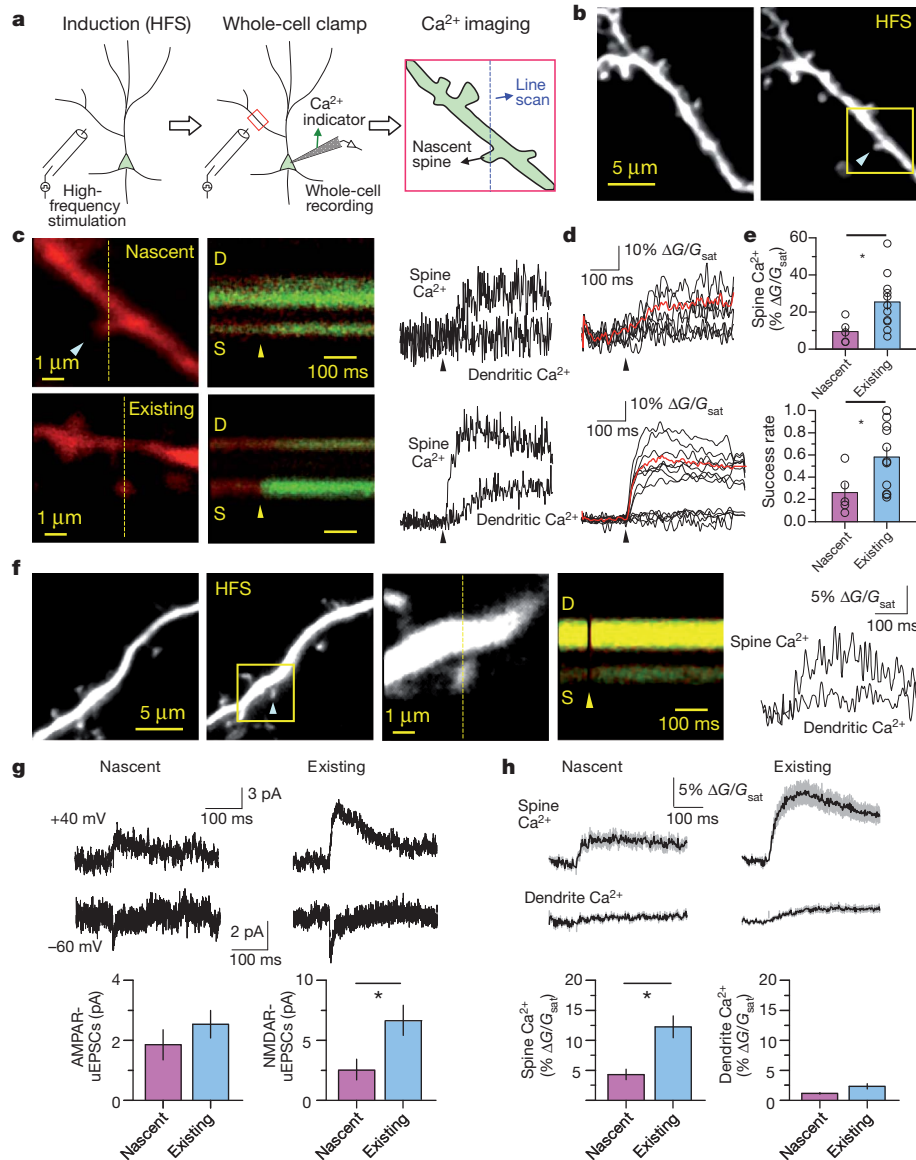


Figure 4 | Functional characterization of new spines. **a**, Schematic of the experimental procedure. HFS was delivered approximately 30 μm from the target dendritic region (red box). After nascent spines were identified, Ca^{2+} indicator was loaded into the cell through a whole-cell recording pipette. The newly generated spine was examined at higher temporal and spatial resolution to measure synaptically evoked Ca^{2+} transients in the spine head and perform optical quantal analysis. **b**, Images before and after HFS showing the new spine (arrowhead) and the area subsequently analysed at higher resolution (yellow box). **c**, Images (left) of newly generated (top) and pre-existing (bottom) spines filled with the fluorophore Alexa 594 (red, 20 μM) and Fluo-5F (green, 300 μM). Fluorescence was collected (middle) and quantified (right) from a line-scan intersecting the spine (S) and dendrite (D) after electrical stimulation (arrowhead). The increases in green fluorescence indicate Ca^{2+} entry. **d**, Green fluorescence transients collected in consecutive trials (black) showing successes and failures. The average 'success' fluorescence transient is also shown (red).

The AMPA receptor (AMPA) and NMDAR content of new spines was characterized using glutamate uncaging after HFS-induced spinogenesis. In all cases, we detected fast uEPSCs at a holding potential of -60 mV and large prolonged uEPSCs at $+40$ mV (Fig. 4g), consistent with AMPAR- and NMDAR-mediated currents previously characterized in these cells²⁵. Similarly, uncaging-evoked Ca^{2+} transients were clearly visible at -60 mV (Fig. 4h) but were larger at $+40$ mV (data not shown), consistent with the known properties of NMDAR-mediated Ca^{2+} influx. Despite similar-sized AMPAR-uEPSCs in new and pre-existing spines, NMDAR-uEPSCs and Ca^{2+} influx were significantly lower in nascent

e, Average amplitude (top, $\Delta\text{G}/\text{G}_{\text{sat}}$) and rate (bottom) of success trials in nascent and neighbouring existing spines for neurons held at -60 mV (nascent and existing: spine Ca^{2+} $\Delta\text{G}/\text{G}_{\text{sat}}$: $9.45 \pm 2.8\%$, $n = 5$, $25.5 \pm 4.4\%$, $n = 11$, $P < 0.05$; success rate: 0.26 ± 0.09 , $n = 5$, $0.58 \pm 0.09\%$, $n = 11$, $P < 0.05$). **f**, Similar experiments as those in **a–c** using glutamate uncaging to characterize the glutamate receptors on the nascent spine. **g**, Examples (top) and average amplitudes (bottom) of AMPAR- and NMDAR-mediated uEPSCs at holding potentials of -60 and $+40$ mV, respectively, using 1 ms uncaging pulses (nascent and existing: AMPAR-uEPSC: -1.9 ± 1.4 pA, $n = 8$, -2.5 ± 0.5 pA, $n = 18$, $P > 0.05$; NMDAR-uEPSC: 2.5 ± 1.4 pA, $n = 8$, 6.3 ± 1.3 pA, $n = 18$, $P < 0.05$). **h**, Average Ca^{2+} transients measured in spine heads and dendrites for neurons held at -60 mV (nascent and existing: spine Ca^{2+} $\Delta\text{G}/\text{G}_{\text{sat}}$: $4.3 \pm 0.9\%$, $n = 8$, $12.3 \pm 1.8\%$, $n = 18$, $P < 0.05$; dendrite Ca^{2+} $\Delta\text{G}/\text{G}_{\text{sat}}$: $1.1 \pm 0.1\%$, $n = 8$, $2.3 \pm 0.4\%$, $n = 18$, $P > 0.05$). Error bars, s.e.m.

than pre-existing spines (Fig. 4g, h), similar to previous descriptions of spontaneously appearing new spines in hippocampal organotypic slices²⁶. Therefore the smaller Ca^{2+} transients measured in nascent spines by synaptic activation probably reflect a smaller number of postsynaptic NMDARs.

In this study, we established a protocol for the reliable and spatio-temporally precise induction of spinogenesis. These experiments demonstrate that glutamate is sufficient to trigger rapid spine formation and suggest that neurons use glutamate release to establish circuit wiring. Thus these data support the hypothesis that axonal growth and

glutamate release may be the triggering event in synapse formation such that axonal bouton localization is an important early step for precise neuronal circuit formation^{10,27}. Given the involvement of NMDARs, Ca²⁺ stores, cAMP, PKA and MAPK in activity-dependent spinogenesis, it is likely that many neuromodulators that regulate these molecules may influence the capacity or threshold for new spine formation. For instance, activation of dopaminergic, serotonergic or adrenergic receptors that signal by G α_s may facilitate spinogenesis, whereas receptors that activate G α_i -coupled signalling may function as inhibitory signals.

Lastly, we provide experimental evidence that spines that grow *de novo* in developing cortical tissue become rapidly functionally integrated into the circuit such that they sense synaptically released glutamate through AMPARs and NMDARs. Whether these nascent spines are rapidly physically associated with a presynaptic bouton and display the ultrastructural correlates of a synapse is unknown^{26,28–30}. Our results indicate that spines can grow *de novo* without the need for a filopodial intermediate and probably without a dendritic-shaft synapse stage. In total, this study demonstrates that synaptic activity can rapidly modify neuronal connectivity with high accuracy by generating new circuit elements.

METHODS SUMMARY

All procedures on animals followed protocols approved by the Harvard Standing Committee on Animal Care and National Institutes of Health guidelines. *In utero* electroporation of EGFP was performed at embryonic day 15.5 in C57BL/6 mice. All studies were performed on layer 2/3 pyramidal neurons in acute coronal slices identified by their characteristic morphology, position in the slice and expression of GFP. Two-photon glutamate uncaging and imaging was performed using custom microscopes. To induce spine growth, 40 uncaging pulses were delivered at varying frequencies to a spot approximately 0.5 μ m from the edge of the dendrite. Synaptically induced spine growth was triggered with two high-frequency stimuli (100 pulses at 100 Hz) separated by 10 s delivered by a bipolar electrode positioned approximately 30 μ m from the target dendrite. For Ca²⁺ imaging, neurons were loaded through the whole-cell recording electrode with Alexa Fluor-594 (20 μ M) and Fluo-5F (300 μ M) and the amplitudes of fluorescence transients were quantified as a fraction of the maximal green fluorescence achieved in saturating (sat) Ca²⁺ concentrations (G_{sat}). For optical quantal analysis, the synapse associated with a visualized spine was stimulated using a closely positioned glass electrode. The position of the electrode and stimulus intensity were adjusted until (1) Ca²⁺ transients were evoked in the spine head that demonstrated stochastic failures and successes, and (2) Ca²⁺ transients in other spines and the dendritic shaft in the field of view were not evoked. Fisher's exact test was used to compare the efficacy of spinogenesis across conditions. For each spinogenesis trial, an observer blind to the experimental condition was asked to identify if (1) a new spine had grown and (2), if so, how many spines had grown.

Received 25 October 2010; accepted 4 March 2011.

Published online 8 May 2011.

- Grutzendler, J., Kasthuri, N. & Gan, W. B. Long-term dendritic spine stability in the adult cortex. *Nature* **420**, 812–816 (2002).
- Hofer, S. B., Mrsic-Flogel, T. D., Bonhoeffer, T. & Hubener, M. Experience leaves a lasting structural trace in cortical circuits. *Nature* **457**, 313–317 (2009).
- Trachtenberg, J. T. *et al.* Long-term *in vivo* imaging of experience-dependent synaptic plasticity in adult cortex. *Nature* **420**, 788–794 (2002).
- Zuo, Y., Yang, G., Kwon, E. & Gan, W. B. Long-term sensory deprivation prevents dendritic spine loss in primary somatosensory cortex. *Nature* **436**, 261–265 (2005).
- Rakic, P., Bourgeois, J. P., Eckenhoff, M. F., Zecevic, N. & Goldman-Rakic, P. S. Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. *Science* **232**, 232–235 (1986).
- Dalva, M. B., McClelland, A. C. & Kayser, M. S. Cell adhesion molecules: signalling functions at the synapse. *Nature Rev. Neurosci.* **8**, 206–220 (2007).

- Scheiffele, P. Cell-cell signaling during synapse formation in the CNS. *Annu. Rev. Neurosci.* **26**, 485–508 (2003).
- Sudhof, T. C. Neuroligins and neuroligins link synaptic function to cognitive disease. *Nature* **455**, 903–911 (2008).
- Yuste, R. & Bonhoeffer, T. Genesis of dendritic spines: insights from ultrastructural and imaging studies. *Nature Rev. Neurosci.* **5**, 24–34 (2004).
- Miller, M. & Peters, A. Maturation of rat visual cortex. II. A combined Golgi-electron microscope study of pyramidal neurons. *J. Comp. Neurol.* **203**, 555–573 (1981).
- Richards, D. A. *et al.* Glutamate induces the rapid formation of spine head protrusions in hippocampal slice cultures. *Proc. Natl Acad. Sci. USA* **102**, 6166–6171 (2005).
- Fiala, J. C., Feinberg, M., Popov, V. & Harris, K. M. Synaptogenesis via dendritic filopodia in developing hippocampal area CA1. *J. Neurosci.* **18**, 8900–8911 (1998).
- Sorra, K. E., Fiala, J. C. & Harris, K. M. Critical assessment of the involvement of perforations, spinules, and spine branching in hippocampal synapse formation. *J. Comp. Neurol.* **398**, 225–240 (1998).
- Engert, F. & Bonhoeffer, T. Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature* **399**, 66–70 (1999).
- Maletic-Savatic, M., Malinow, R. & Svoboda, K. Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity. *Science* **283**, 1923–1927 (1999).
- Harvey, C. D. & Svoboda, K. Locally dynamic synaptic learning rules in pyramidal neuron dendrites. *Nature* **450**, 1195–1200 (2007).
- Harvey, C. D., Yasuda, R., Zhong, H. & Svoboda, K. The spread of Ras activity triggered by activation of a single dendritic spine. *Science* **321**, 136–140 (2008).
- Matsuzaki, M., Honkura, N., Ellis-Davies, G. C. & Kasai, H. Structural basis of long-term potentiation in single dendritic spines. *Nature* **429**, 761–766 (2004).
- Tanaka, J. *et al.* Protein synthesis and neurotrophin-dependent structural plasticity of single dendritic spines. *Science* **319**, 1683–1687 (2008).
- Yasuda, H., Barth, A. L., Stellwagen, D. & Malenka, R. C. A developmental switch in the signaling cascades for LTP induction. *Nature Neurosci.* **6**, 15–16 (2003).
- Zaccolo, M. & Pozzan, T. Discrete microdomains with high concentration of cAMP in stimulated rat neonatal cardiac myocytes. *Science* **295**, 1711–1715 (2002).
- Di Cristo, G. *et al.* Requirement of ERK activation for visual cortical plasticity. *Science* **292**, 2337–2340 (2001).
- Chalifoux, J. R. & Carter, A. G. GABA_B receptors modulate NMDA receptor calcium signals in dendritic spines. *Neuron* **66**, 101–113 (2010).
- Skeberdis, V. A. *et al.* Protein kinase A regulates calcium permeability of NMDA receptors. *Nature Neurosci.* **9**, 501–510 (2006).
- Busetto, G., Higley, M. J. & Sabatini, B. L. Developmental presence and disappearance of postsynaptically silent synapses on dendritic spines of rat layer 2/3 pyramidal neurons. *J. Physiol. (Lond.)* **586**, 1519–1527 (2008).
- Zito, K., Scheuss, V., Knott, G., Hill, T. & Svoboda, K. Rapid functional maturation of nascent dendritic spines. *Neuron* **61**, 247–258 (2009).
- Hamori, J. The inductive role of presynaptic axons in the development of postsynaptic spines. *Brain Res.* **62**, 337–344 (1973).
- De Roo, M., Klausner, P., Mendez, P., Poglia, L. & Muller, D. Activity-dependent PSD formation and stabilization of newly formed spines in hippocampal slice cultures. *Cereb. Cortex* **18**, 151–161 (2008).
- Nagerl, U. V., Kostinger, G., Anderson, J. C., Martin, K. A. & Bonhoeffer, T. Protracted synaptogenesis after activity-dependent spinogenesis in hippocampal neurons. *J. Neurosci.* **27**, 8149–8156 (2007).
- Knott, G. W., Holtmaat, A., Wilbrecht, L., Welker, E. & Svoboda, K. Spine growth precedes synapse formation in the adult neocortex *in vivo*. *Nature Neurosci.* **9**, 1117–1124 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Sabatini laboratory for their comments on the manuscript and assistance with data analysis. We are grateful to S. Nazia for technical support and for acting as the blind evaluator. This work was supported by a SFARI grant from the Simons Foundation and the National Institute of Neurological Disorders and Stroke (NS046579).

Author Contributions H.B.K. and B.L.S. designed the experiments and wrote the paper. H.B.K. performed all the experiments, analysed the data (other than spine counting by a blind, third-party observer) and prepared the figures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.L.S. (bsabatini@hms.harvard.edu).

Earliest evidence of mammalian social behaviour in the basal Tertiary of Bolivia

Sandrine Ladevèze¹, Christian de Muizon², Robin M. D. Beck³, Damien Germain² & Ricardo Cespedes-Paz⁴

The vast majority of Mesozoic and early Cenozoic metatherian mammals (extinct relatives of modern marsupials) are known only from partial jaws or isolated teeth, which give insight into their probable diets and phylogenetic relationships but little else. The few skulls known are generally crushed, incomplete or both^{1–4}, and associated postcranial material is extremely rare. Here we report the discovery of an exceptionally large number of almost undistorted, nearly complete skulls and skeletons of a stem-metatherian, *Pucadelphys andinus*, in the early Palaeocene epoch⁵ of Tiupampa in Bolivia^{6–8}. These give an unprecedented glimpse into early metatherian morphology, evolutionary relationships and, especially, ecology. The remains of 35 individuals have been collected, with 22 of these represented by nearly complete skulls and associated postcrania. These individuals were probably buried in a single catastrophic event, and so almost certainly belong to the same population⁹. The preservation of multiple adult, sub-adult and juvenile individuals in close proximity (<1 m²) is indicative of gregarious social behaviour or at least a high degree of social tolerance and frequent interaction. Such behaviour is unknown in living didelphids, which are highly solitary and have been regarded, perhaps wrongly, as the most generalized living marsupials. The Tiupampa *P. andinus* population also exhibits strong sexual dimorphism, which, in combination with gregariousness, suggests strong male–male competition and polygyny. Our study shows that social interactions occurred in metatherians as early as the basal Palaeocene and that solitary behaviour may not be plesiomorphic for Metatheria as a whole.

The locality of Tiupampa, Bolivia^{10,11}, has yielded an abundant and remarkably preserved mammal fauna from the earliest Palaeocene⁶. Some of the most spectacular fossils are numerous partial skulls and skeletons of the small (about 20 g) stem-metatherian, *P. andinus* (Fig. 1 and Supplementary Information). In 1985, four partial skeletons (two of them with a skull) and eight additional sub-complete to partial skulls were discovered in an area of less than a square metre. Some isolated jaws and teeth were also collected in this area. In 1997, a second concentration of *P. andinus* remains was discovered approximately 3 m away from the first area, with 23 skulls and associated postcranial skeletons present in approximately a square metre (Fig. 1). These previously undescribed 1997 specimens have now been prepared and, together with the 1985 specimens, provide significant new information about the ecology of *Pucadelphys*.

In both areas, the skeletons were often intermixed with little or no disarticulation. Such preservation demonstrates that there was no postmortem transportation. The abundance of sub-complete articulated skeletons of other small vertebrates (pantodonts, two other metatherians, frogs, lizards, snakes, newborn crocodiles) and the presence of numerous crocodile eggs (including an intact nest) all within a very restricted area (less than 100 m²) suggests that a single catastrophic event buried this fauna *in situ*; sedimentological evidence indicates that the fauna existed by a pond or ox-bow lake, adjacent

to the bank of a large meandering river^{9,10,12}. *P. andinus* appears to have been mainly terrestrial with arboreal abilities¹³, and probably lived in burrows¹⁰. In this context, it is highly probable that the *P. andinus* individuals were living together and therefore comprise a sample of the same population. Of the 35 specimens represented by partial skulls and/or partial skeletons, 22 present exceptionally preserved skulls and mandibles (12 adult females, six adult males, four sub-adults), which is unprecedented in the fossil record of pre-Neogene therians. Indeed, *Pucadelphys* is now known from more specimens than several living marsupials (for example, *Glironia venusta*, *Microperoryctes murina*).

Four of the skulls of the *P. andinus* sample represent sub-adults, given that their adult dentition is not fully erupted: P3 (third upper premolar) and m4 (fourth lower molar) are both partially erupted, while tooth M4 is unerupted. From comparison with extant didelphid marsupials, this dental stage may correspond to either immature or sexually precocious individuals^{14,15}. A single lower jaw¹⁶ retains an alveolus for tooth dp3 (third deciduous lower premolar), and tooth m3 is only partially erupted, suggesting that it is from an immature individual^{14,15}. Of the 18 skulls with fully erupted adult dentition, six are considerably larger than the others and exhibit longer canines and higher sagittal and occipital crests. However, the size, relative proportions and morphology of the postcanine dentition indicate that all 18 individuals belong to a single species. The morphological differences are here interpreted as evidence of strong sexual dimorphism (Fig. 2).

Body-mass estimates (Supplementary Information) indicate that the larger skulls, considered here to represent males, are from individuals that weighed on average 35% more than the smaller-skulled (female) individuals (Supplementary Information). Strong sexual size dimorphism is found in some extant marsupials that are probably ecologically similar to *P. andinus* in terms of body size, diet (based on dental morphology) and pattern of locomotion (terrestrial–scansorial to arboreal), namely small South American didelphids (for example, species of *Gracilinanus*, *Marmosa*, *Marmosops* and *Monodelphis*¹⁷) and small Australian dasyurids (for example, *Antechinus*¹⁸). However, the dimorphism seen in these living taxa appears to be connected with semelparity (all males die shortly after a single breeding season)^{17,19}, and *P. andinus* was clearly not semelparous, given the coexistence of adult males with sub-adults and juveniles. Sexual size dimorphism in small dasyurids may be related to male dominance over females, which are generally territorial and commonly aggressive towards conspecifics²⁰. However, the presence of multiple adult males and females in the *P. andinus* sample indicates that females were non-territorial, and so the observed sexual size dimorphism is more likely to be connected with male–male competition, and hence polygyny (one male mates with more than one female).

Principal component analyses (PCAs) based on a maximum of 43 craniodental measurements reveal good discrimination of putative male and female individuals, as well as between adults and sub-adults (Fig. 3, Supplementary Information). A sexual dimorphism index (SDI,

¹Department of Paleontology, Royal Belgian Institute of Natural Sciences, 29 rue Vautier, B-1000 Brussels, Belgium. ²Département Histoire de la Terre UMR7207 CR2P (CNRS, MNHN, UPMC), 8 rue Buffon, Muséum national d'Histoire naturelle de Paris, 8 rue Buffon, F-75005 Paris, France. ³Department of Mammalogy, American Museum of Natural History, Central Park West 79th Street, New York, New York 10024, USA. ⁴Museo de Historia Natural Alcide D'Orbigny, Av. Potosí N-1458, Cochabamba, Bolivia.

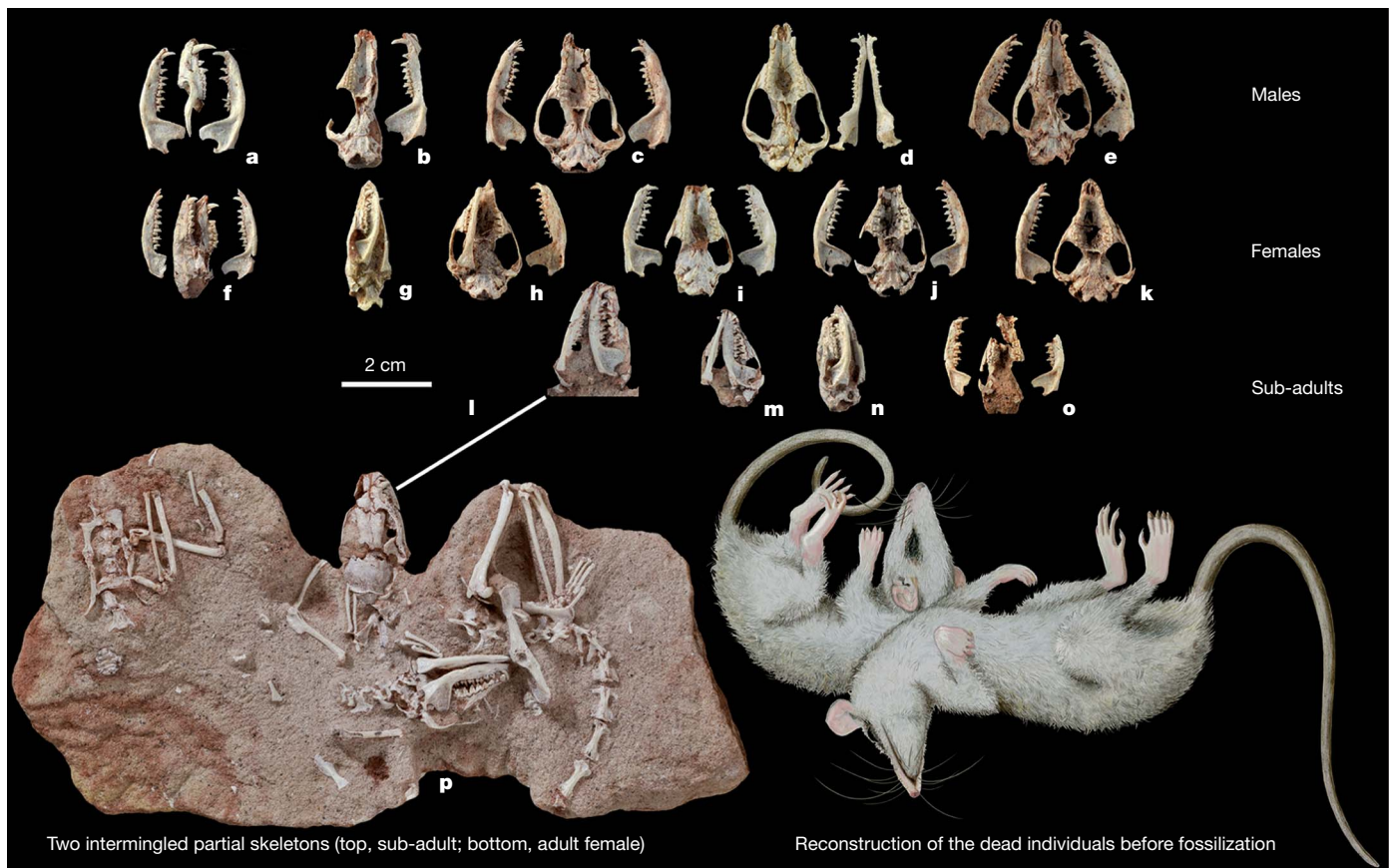


Figure 1 | Skulls and partial skeletons of *P. andinus* from the same population of basal Palaeocene beds at Tiupampa (Bolivia). a, MHNC 8364. b, MHNC 8377. c, MHNC 8381. d, MHNC 8382. e, MHNC 8266. f, MHNC 8386. g, MHNC 8389. h, MHNC 8379. i, MHNC 8376. j, MHNC 8380. k, MHNC 8378. l, MHNC 8391. m, MHNC 8388. n, MHNC 8265. o, MHNC

8384. p, MHNC 8392. The differences of position in skulls and bones between the block with tangled skeletons (left) and the reconstruction (right) is interpreted as bone displacement after flesh decomposition and prior to sediment consolidation.



as defined in the Supplementary Information), together with identification of the most dimorphic parameters by analyses of variance (ANOVAs), reveals that *P. andinus* males are characterized by longer and wider skulls (markedly expanded posteriorly), longer and higher dentaries and larger canines. These sexually dimorphic craniodental features have been shown to be reliable indicators of social structure in living faunivorous mammals²¹, and further supports the existence of strong male–male competition and polygyny in *P. andinus*.

Beyond these sexually dimorphic morphological features, the *P. andinus* sample also reveals extreme intra-specific and intra-individual variability in molar morphology that is unrelated to sexual dimorphism. Indeed, if found as isolated elements, several teeth would probably have been identified as representing different taxa. The most variable molar features are overall robustness and size, morphology of the styler cusps, protocone width, talonid labial angle, entoconid size, paraconid position and, most strikingly, the overall morphology and proportion of the ultimate molars (Fig. 4). Such intra-population variation in molar morphology has not been observed in similar-sized modern didelphids (except for the styler cusp variation²²), but is present in some populations of the much larger *Didelphis virginiana*.

The population structure seen in the *Pucadelphys* fossil assemblage comprises multiple adults of both sexes and different ages (based on the degree of tooth wear), together with sub-adults and one juvenile. Extant didelphids are solitary animals occurring at low overall population densities and are usually highly intolerant of conspecifics²³.

Figure 2 | Skulls and mandibles of male (right, MHNC 8266) and female (left, MHNC 8378) of *P. andinus*. Top, dorsal view; middle, ventral view; bottom, lateral view.

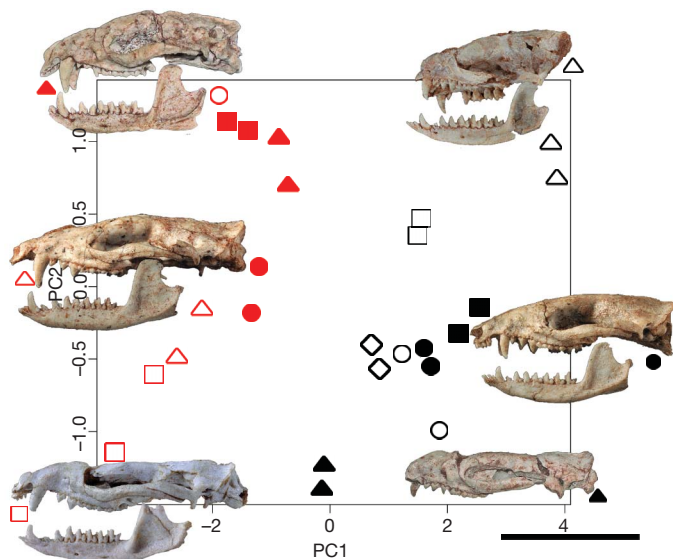


Figure 3 | PCA of 32 cranial and dental variables in 13 adult specimens of *P. andinus*. (This is PCA3; see Supplementary Information for the method and other PCAs). The repartition of the variables on the two first axes of the PCA highlights a good discrimination of the males (left, in red) and females (right, in black). The symbols refer to specimens: open black circle, MHNC 8376; open red circle, MHNC 8377; filled black circle, MHNC 8378; open black square, MHNC 8379; filled black square, MHNC 8380; filled red circle, MHNC 8381; open red square, MHNC 8382; filled red square, MHNC 8364; open red triangle, MHNC 8266; open black triangle, MHNC 8386; filled red triangle, YPFB Pal 6105; filled black triangle, YPFB Pal 6108; open black diamond, YPFB Pal 6110. Scale, 2 cm.

Female didelphids are highly territorial and their home ranges rarely overlap. Most dasyurids are also solitary²⁴. By contrast, the presence of the fossils of at least 35 individuals, including 12 adult females, within an area of a few square metres strongly suggests that *P. andinus* was gregarious and that females were non-territorial.

Available data suggest that didelphids have a non-monogamous, probably promiscuous, mating system, with males actively searching

for females during the breeding season²⁵. Even during the breeding season, no didelphid species has been reported to form groups as large as those inferred in *P. andinus*. Didelphids never live in groups that include adults of both sexes together with sub-adults and juveniles, again unlike *P. andinus*. Dasyurid mating systems are either based on promiscuous matings between solitary individuals in overlapping home ranges²³ or lek promiscuity, with individual females visiting aggregations of males²⁶. The presence of multiple sub-adults and at least one juvenile in the *P. andinus* sample, together with the female-biased sex ratio of 12 adult females and six adult males, suggests that the inferred gregariousness in this species is not connected to mating.

Although most dasyurids are predominantly or entirely solitary, nesting groups occur in some species. These usually comprise a mother with young and postweaning juveniles, but nesting groups of between two and 40 individuals have been observed²⁶. However, such large nesting groups have only been reported in southern Australia, and communal nesting only occurs during winter with the onset of very cold nights²⁶. This pattern, together with the observation that the dasyurid *Sminthopsis crassicaudata* also forms winter nests with the house mouse *Mus musculus* (normally its prey)²⁷, suggests that the primary purpose of large nesting groups is thermoregulatory, with huddling allowing energetic savings. Bolivia in the early Cenozoic was approximately at the same latitude as it is now (the present latitude of Tiupampa is about 18° S) but at much lower altitude than today. The climate at Tiupampa was probably similar to that of the modern Amazonian rainforest, namely tropical with temperatures well above freezing throughout the year and with wet–dry seasonality, which is reinforced by the fact that the neotropical temperatures during the Palaeocene were significantly warmer than they are now (about +6 °C warmer)²⁸. The gregariousness inferred in *P. andinus* was therefore unlikely to be connected with thermoregulation.

In summary, the fossils discussed here demonstrate that the early Palaeocene *P. andinus* population at Tiupampa was a socially interacting group, characterized by gregariousness, male–male competition and polygyny. The probable tropical climate of Tiupampa during the early Palaeocene suggests that breeding in *P. andinus* was probably aseasonal, which may have encouraged male–male competition and hence led to the evolution of sexual size and shape dimorphism. The discovery at Tiupampa of six articulated and intermingled skeletons of another metatherian taxon, *Andinodelphys cochabambensis* (which was probably more arboreal than *Pucadelphys*)¹³, may indicate gregariousness in this taxon as well. The Tiupampa metatherians therefore indicate for the first time that, as early as the basal Palaeocene, gregarious social behaviour was present in metatherians, and that the highly solitary behaviour of living didelphids may be a derived rather than primitive characteristic.

METHODS SUMMARY

The detailed method and calculations for estimating the body mass of *P. andinus* is provided in the Supplementary Information. Multivariate analyses were performed (PCAs and ANOVAs) and are described in the Supplementary Information. The quantification of the sexual dimorphism was possible thanks to a new index, the description of which is given in the Supplementary Information. Also see Supplementary Information for detailed Methods, data sets, measurements, calculations and further references.

Received 17 January; accepted 3 March 2011.

Published online 8 May 2011.

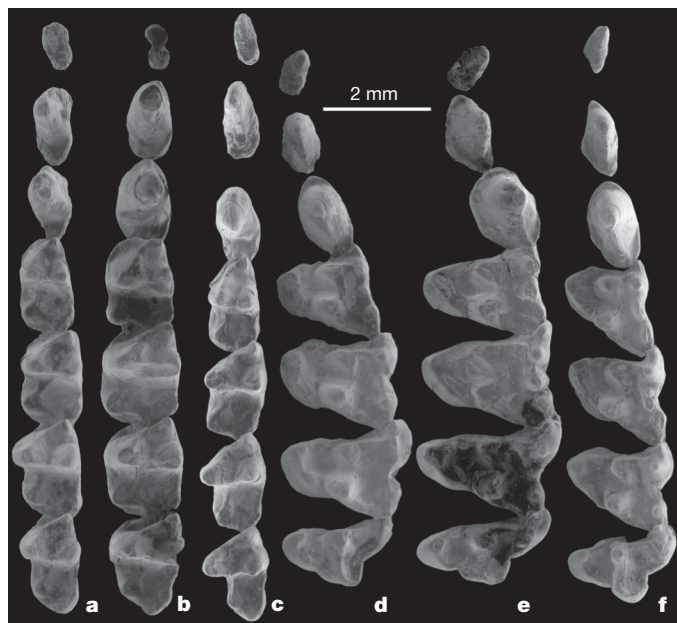


Figure 4 | Upper and lower dental series of *P. andinus* to show the individual variation of dental morphology. **a, d**, Left lower (**a**) and left upper (**d**) tooth row of MHNC 8378 (female). **b, e**, Right (reversed) lower (**b**) and right (reversed) upper (**e**) tooth row of MHNC 8381 (male). **c, f**, Right (reversed) lower (**c**) and right (reversed) upper (**f**) tooth row of MHNC 8266 (male).

1. Luo, Z.-X., Ji, Q., Wible, J. R. & Yuan, C.-X. An early Cretaceous tribosphenic mammal and metatherian evolution. *Science* **302**, 1934–1940 (2003).
2. Rougier, G. W., Wible, J. R. & Novacek, M. J. Implications of *Deltatheridium* specimens for early marsupial history. *Nature* **396**, 459–463 (1998).
3. Szalay, F. S. & Trofimov, B. The Mongolian Late Cretaceous *Asiatherium*, and the early phylogeny and paleobiogeography of Metatheria. *J. Vert. Pal.* **16**, 474–509 (1996).
4. Horowitz, I. et al. Cranial anatomy of the earliest marsupials and the origin of opossums. *PLoS ONE* **4**, e8278 (2009).

5. Gelfo, J. N., Goin, F. J., Woodburne, M. O. & de Muizon, C. Biochronological relationships of the earliest South American Paleogene mammalian faunas. *Palaeontology* **52**, 251–269 (2009).
6. Marshall, L. G. & de Muizon, C. in *Pucadelphys andinus* (Marsupialia, Mammalia) from the early Paleocene of Bolivia (ed. de Muizon, C.) Vol. 165, 21–90 (Mémoires du Muséum national d'Histoire naturelle, 1995).
7. de Muizon, C. *Mayulestes ferox*, a boryanoid (Metatheria, Mammalia) from the early Palaeocene of Bolivia. Phylogenetic and palaeobiologic implications. *Geodiversitas* **20**, 19–142 (1998).
8. de Muizon, C., Cifelli, R. L. & Céspedes Paz, R. The origin of the dog-like borhyaenoid marsupials of South America. *Nature* **389**, 486–489 (1997).
9. de Muizon, C. & Céspedes, R. in *Origins and Evolution of Cenozoic South American Mammals* (eds Rosenberger, A. & Tejedor, M.) (Springer, in the press).
10. Marshall, L. G., de Muizon, C. & Sigogneau-Russel, D. *Pucadelphys andinus* (Marsupialia, Mammalia) from the Early Paleocene of Bolivia (Mémoires du Muséum national d'Histoire naturelle, 1995).
11. de Muizon, C. in *Fósiles y Fácies de Bolivia* Vol. I Vertebrados Vol. 12 (3–4) (ed. Suárez-Soruco, R.) 575–624 (Revista Técnica, Yacimientos Petrolíferos Fiscales Bolivianos, 1992).
12. Marshall, L. G., Sempere, T. & Butler, R. F. Chronostratigraphy of the mammal-bearing Paleocene of South America. *J. S. Am. Earth Sci.* **10**, 49–70 (1997).
13. de Muizon, C. & Argot, C. in *Predators with Pouches: the Biology of Carnivorous Marsupials* (eds Jones, M., Dickman, C. & Archer, M.) 42–63 (Surrey Beatty & Sons, 2003).
14. Atramontowicz, M. Dynamique de population chez trois marsupiaux didelphidés de Guyane. *Biotropica* **18**, 136–149 (1986).
15. Tyndale-Biscoe, C. H. & MacKenzie, R. B. Reproduction in *Didelphis marsupialis* and *Didelphis albiventris* in Colombia. *J. Mamm.* **57**, 249–265 (1976).
16. Cifelli, R. L. & de Muizon, C. Tooth eruption and replacement pattern in early marsupials. *C. R. Acad. Sci. Ser. D* **326**, 215–220 (1998).
17. Astúa, D. Cranial sexual dimorphism in New World marsupials and a test of Rensch's rule in Didelphidae. *J. Mamm.* **91**, 1011–1024 (2010).
18. Fisher, D. O. & Cockburn, A. The large-male advantage in brown antechinuses: female choice, male dominance, and delayed male death. *Behav. Ecol.* **17**, 164–171 (2005).
19. Kraaijeveld, K., Kraaijeveld-Smit, F. J. L. & Adcock, G. J. Does female mortality drive male semelparity in dasyurid marsupials? *Proc. R. Soc. Lond. B* **270** (Suppl. 2), S251–S253 (2003).
20. Soderquist, T. Spatial organization of the arboreal carnivorous marsupial, *Phascogale tapoatafa*. *J. Zool.* **237**, 385–398 (1995).
21. Plavcan, J. M. Inferring social behavior from sexual dimorphism in the fossil record. *J. Hum. Evol.* **39**, 327–344 (2000).
22. Hershkovitz, P. Composition of the family Didelphidae Gray, 1821 (Didelphoidea: Marsupialia), with a review of the morphology and behavior of the included four-eyed pouched opossums of the genus *Philander* Tiedemann, 1808. *Feldiana Zool.* **86**, 1–103 (1997).
23. Russell, E. M. Social behaviour and social organization of marsupials. *Mammal Rev.* **14**, 101–154 (1984).
24. Oakwood, M. Spatial and social organization of a carnivorous marsupial *Dasyurus hallucatus* (Marsupialia: Dasyuridae). *J. Zool.* **257**, 237–248 (2002).
25. Cáceres, N. C., Napoli, R. P., Lopes, W. H., Casella, J. & Gazeta, G. S. Natural history of the marsupial *Thylamys macrurus* (Mammalia, Didelphidae) in fragments of savannah in southwestern Brazil. *J. Nat. Hist.* **41**, 1979–1988 (2007).
26. Lazenby-Cohen, K. A. & Cockburn, A. Lek promiscuity in a semelparous mammal, *Antechinus stuartii* (Marsupialia: Dasyuridae)? *Behav. Ecol. Sociobiol.* **22**, 195–202 (1988).
27. Morton, S. R. Torpor and nest-sharing in free-living *Sminthopsis crassicaudata* (Marsupialia) and *Mus musculus* (Rodentia). *J. Mamm.* **59**, 569–575 (1978).
28. Head, J. J. *et al.* Giant boid snake from the Palaeocene neotropics reveals hotter past equatorial temperatures. *Nature* **457**, 715–717 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Field expeditions were supported by the National Geographic Society (grants 2908-84) in 1985 and by the 'Institut Français d'Études Andines' in 1997. The specimens collected are the property of the 'Museo de Historia Natural Alcide d'Orbigny' in Cochabamba, Bolivia. Additional financial support was provided by Research Project MO/36/020 of the Belgian Federal Science Policy Office (S.L.), and NSF grant DEB-0743039 (R.M.D.B.), in collaboration with R. Voss at the American Museum of Natural History. Photographs are by C. Lemzaouda and P. Loubry, scanning electron microscope photographs are by J. Cillis and the reconstruction in Fig. 1 is by S. Fernandez. We thank E. Westwig for access to the AMNH collections, and O. Lambert, A. Pradel, and T. Soderquist for discussions.

Author Contributions C.d.M. and R.C.-P. collected the specimens at Tiupampa. C.d.M. prepared the specimens. S.L. and C.d.M. initiated and organized the project. S.L., C.d.M. and R.M.D.B. wrote the paper. S.L., C.d.M., R.M.D.B., D.G. and R.C.-P. discussed the results and commented on the manuscript at all stages. Measurements were taken by S.L. and R.M.D.B. PCAs and ANOVAs were performed by S.L. and D.G.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.L. (sandrine.ladeveze@naturalsciences.be) or C.M. (muizon@mnhn.fr).

Inferring nonlinear mantle rheology from the shape of the Hawaiian swell

N. Asaadi¹, N. M. Ribe² & F. Sobouti¹

The convective circulation generated within the Earth's mantle by buoyancy forces of thermal and compositional origin is intimately controlled by the rheology of the rocks that compose it. These can deform either by the diffusion of point defects (diffusion creep, with a linear relationship between strain rate and stress) or by the movement of intracrystalline dislocations (nonlinear dislocation creep)^{1,2}. However, there is still no reliable map showing where in the mantle each of these mechanisms is dominant, and so it is important to identify regions where the operative mechanism can be inferred directly from surface geophysical observations. Here we identify a new observable quantity—the rate of downstream decay of the anomalous seafloor topography (swell) produced by a mantle plume—which depends only on the value of the exponent in the strain rate versus stress relationship that defines the difference between diffusion and dislocation creep. Comparison of the Hawaiian swell topography with the predictions of a simple fluid mechanical model shows that the swell shape is poorly explained by diffusion creep, and requires a dislocation creep rheology. The rheology predicted by the model is reasonably consistent with laboratory deformation data for both olivine³ and clinopyroxene⁴, suggesting that the source of Hawaiian lavas could contain either or both of these components.

Three distinct approaches have been used to constrain the rheological structure of the Earth's mantle. The most versatile approach is laboratory deformation experiments wherein the relation between the strain rate $\dot{\epsilon}$ and the deviatoric stress τ in a deforming sample is measured under controlled physical and chemical conditions^{1,2}. Such experiments yield rheological laws of the form $\dot{\epsilon} = D\tau^n$, where n is a power-law exponent and D depends in general on pressure, temperature, grain size and the chemical composition and mineralogy of the sample. Diffusion creep has $n = 1$, whereas $n = 3.5$ for dislocation creep of olivine⁵, the dominant mineral in the uppermost mantle. However, application of these laws to the Earth is subject to large uncertainties, both because several of the properties on which D depends are poorly constrained in the mantle and because geological strain rates are 6–8 orders of magnitude lower than those in the laboratory.

A second, geodynamical approach is to infer the depth variation of mantle viscosity from the rates of surface rebound following deglaciation events⁶, the long-wavelength components of the Earth's nonhydrostatic gravitational equipotential surface (geoid)⁷, or both together⁸. A robust result of this approach is that the lower mantle must be more viscous (by at least a factor of ten) than the upper mantle. However, studies using this approach typically assume a linear rheology ($n = 1$) throughout the mantle, and consequently provide little information about the relative importance of dislocation versus diffusion creep.

The third approach is seismological, and exploits the fact that dislocation creep (but not diffusion creep) causes the crystallographic axes of the crystals in a deforming rock to become aligned in a non-random way⁹. The resulting dependence of elastic wave speed on the propagation direction (seismic anisotropy) can be detected either by its effect on surface waves¹⁰ or by the splitting of shear waves that it

causes¹¹, and provides unambiguous evidence for active dislocation creep somewhere along the path between the source and receiver. However, the precise location of the anisotropic region along this path is difficult to constrain.

Here we show that a more direct determination of the operative deformation mechanism is possible in a well-defined portion of the Earth's uppermost mantle that lies beneath the Hawaiian islands. Since the earliest days of plate tectonic theory, the Hawaiian islands have inspired many fundamental new ideas in global geophysics, including the concept of a hotspot as a fixed locus of mantle melting^{12,13}, the association of hotspots with plumes rising from the lower mantle^{13–15}, the lithospheric flexure model of isostasy¹⁶, and the recognition that hotspots can experience episodes of rapid migration¹⁷. Below we show that Hawaii can also make a novel contribution to our understanding of mantle rheology.

The largest geophysical signature produced by the Hawaiian plume is a broad topographic swell some 1,400 m high and 1,000 km wide (Fig. 1). According to the principle of gravitational equilibrium (isostasy), the downward buoyancy force on the swell (which displaces sea water) must be compensated by an upward buoyancy force provided by low-density material beneath: in this case, the anomalously hot material that ascends in the Hawaiian plume and then spreads laterally over the base of the Pacific plate.

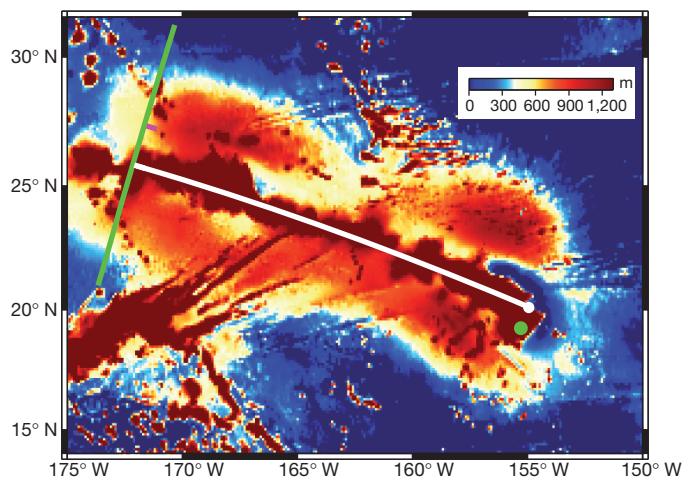


Figure 1 | Residual topography of the Hawaiian swell. This is defined as the observed seafloor topography²⁹ minus the reference topography predicted by a thermal model for cooling oceanic lithosphere³⁰. Negative values of residual topography appear as dark blue, and values exceeding 1,300 m as red-brown. The white line is a segment of a great circle that fits the central axis of the island chain between longitudes 159–173° W. The green dot shows the current location of the centre of the Hawaiian hotspot (Kilauea volcano), and the white dot is its projection onto the central axis. Our analysis considers only the main part of the swell southeast of the green line.

¹Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan 45137-66731, Iran. ²Université Pierre et Marie Curie (Paris 6), Université Paris-Sud, CNRS, Lab FAST, Bâtiment 502, Campus Universitaire, 91405 Orsay, France.

Focusing now on the main part of the swell east of 173° W longitude, we note first that its width is smaller in the middle (around 163° W) than to either side, reflecting a past episode of reduced (by about 20%) upward flux of material in the Hawaiian plume¹⁸. However, the parts of the swell to the left of (older than) and the right of (younger than) the constriction have nearly the same amplitude and width. This crucial observation motivates the following physical model.

We consider the Hawaiian plume as a fixed source of buoyant fluid beneath a plate moving at constant speed^{19,20} (Fig. 2a). The fluid spreads laterally over the plate's base to form a broad plume head with a thickness that is small compared to its lateral dimensions. Because the flow within it is dominantly simple shear on horizontal planes, the viscosity $\eta = \tau/\dot{\epsilon}$ corresponding to the rheology $\dot{\epsilon} = D\tau^n$ is:

$$\eta = D^{-\frac{1}{n}} \frac{1-n}{\dot{\epsilon}} \quad (1)$$

where $\dot{\epsilon} = \left| \frac{\partial \mathbf{u}}{\partial z} \right|$, \mathbf{u} is the horizontal velocity and z is the vertical coordinate.

The thickness of a plume head with the rheology of equation (1) is governed by a partial differential equation that can be derived using the 'lubrication' theory of slow viscous flow in thin layers (Methods). It contains three parameters: the plate speed U , the volumetric supply rate Q of the buoyant material, and the parameter $\sigma = D(g\delta\rho)^n/[(n+2)(n+3)]$ that characterizes its rate of lateral spreading, where g is the gravitational acceleration and $-\delta\rho$ is the density deficit of the

plume relative to the ambient mantle. A scaling analysis of the equation shows that the width of the plume head is proportional to:

$$L_0 = \left(\frac{\sigma Q^{2n+1}}{U^{2n+2}} \right)^{\frac{1}{3n+1}} \quad (2)$$

Conservation of the downstream volume flux then implies that the plume head's thickness is proportional to $h_0 = Q/(UL_0)$. Finally, isostasy implies that the amplitude of the topography is proportional to:

$$S_0 = \frac{Q\delta\rho}{UL_0(\rho_0 - \rho_w)} \quad (3)$$

where $\rho_0 = 3,400 \text{ kg m}^{-3}$ is the density of the ambient mantle and $\rho_w = 1,000 \text{ kg m}^{-3}$ that of sea water.

Figures 2b and c show the steady-state numerical solutions of the lubrication equation for diffusion creep ($n = 1$) and dislocation creep ($n = 3.5$) rheologies. Gravitational spreading is less efficient for a plume head with a dislocation creep rheology, which decays and widens more slowly downstream because the viscosity increases from the centre (where $\dot{\epsilon}$ is largest) towards the edges (where $\dot{\epsilon}$ is smaller). The influence of the rheology is revealed by a similarity solution of the lubrication equation that is valid at distances x far downstream from the hotspot (Methods). It predicts that the swell topography along the axis $y = 0$ should decay downstream as:

$$S \propto S_0 \left(\frac{x - x_0}{L_0} \right)^{-\frac{1}{3n+2}} \quad (4)$$

where x_0 is the virtual position of the hotspot. Accordingly, $S \propto (x - x_0)^{-0.2}$ for diffusion creep and $S \propto (x - x_0)^{-0.08}$ for dislocation creep with $n = 3.5$. The rate of downstream decay of the swell is therefore a sensitive indicator of the rheology of the buoyant material that compensates it.

To compare the predictions of the lubrication model with the observations, we first exclude the parts of the residual topography that are unrelated to the swell (Methods). Using a least-squares procedure, we determine the values of L_0 and S_0 for which the numerical solutions of Figs 2b and c best fit the non-excluded portion ($1.7 \times 10^6 \text{ km}^2$) of the residual topography. These solutions are shown in Fig. 3a (for diffusion creep) and Fig. 3b (for dislocation creep), together with two contours (0 m and 700 m) of the residual topography. The diffusion creep solution does not match well the shape of the swell. It decays too rapidly downstream, and is too narrow near the hotspot and too wide farther downstream. In contrast, the dislocation creep solution matches much better the slow downstream decay of the swell and the shape of its boundary.

The above analysis can be refined by accounting for the variation of the Hawaiian plume flux Q during the past 20 million years. This gave rise to an along-chain variation of the swell's cross-sectional area $A(x)$, which has a minimum around 163° W longitude (Fig. 1). Figure 3 shows the solutions of the time-dependent lubrication equation for $n = 1$ and $n = 3.5$ that provide the best fit to the distribution $A(x)$ determined from the residual topography data¹⁸ (Methods). Again, only the solution with $n = 3.5$ provides an acceptable fit to the residual topography.

We now consider the depth of the low-density material compensating the swell, which is not immediately obvious. An estimate is provided by the geoid-topography ratio (GTR), the ratio of the anomalous height of the gravitational equipotential surface to the topography anomaly. In the limit of a very broad swell, the GTR (in metres of geoid per kilometre of topography) is approximately one-tenth the average depth d (in kilometres) of the low-density compensating material. Early estimates of $\text{GTR} \approx 4\text{--}5 \text{ m km}^{-1}$ for the Hawaiian swell²¹ suggested $d \approx 40\text{--}50 \text{ km}$, above the base of normal oceanic lithosphere of the same age (about 90 million years). However, those estimates were too low, owing to incomplete removal of the effect of the volcanic islands, which are more shallowly compensated (by the plate flexure mechanism) than the swell itself. Refined estimates that correct for this bias²² yield $\text{GTR} \approx 7\text{--}8 \text{ m km}^{-1}$, indicating that the

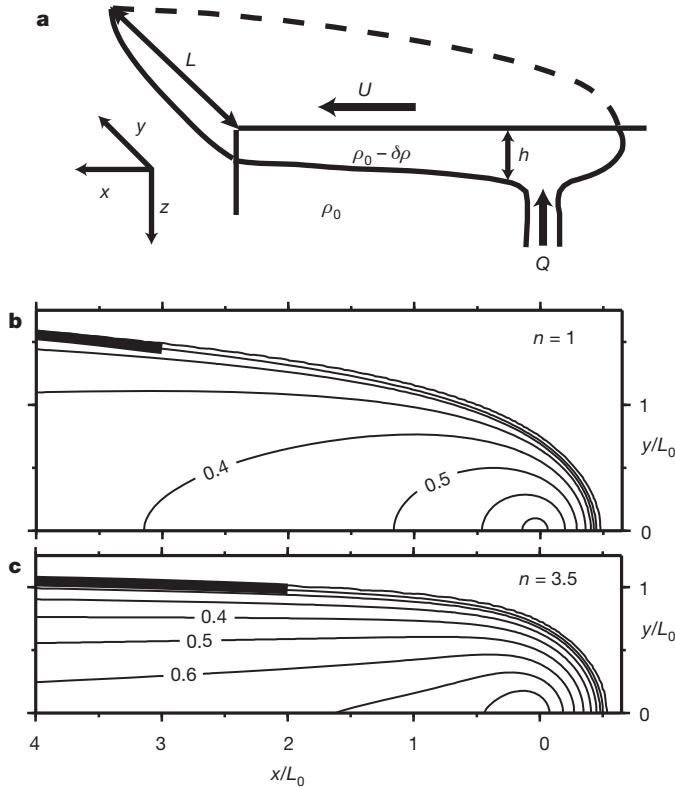


Figure 2 | Lubrication-theory model for the Hawaiian swell. **a**, Model geometry (oblique cut-away view). Buoyant fluid with a density deficit $-\delta\rho$ relative to the ambient mantle and a diffusion creep ($n = 1$) or dislocation creep ($n = 3.5$) rheology is released at a volumetric rate Q beneath a plate moving at speed U . The fluid spreads laterally to form a thin plume head of thickness h and half-width $L(x) \gg h$. **b**, Steady-state plume head thickness $h(x, y)$ predicted by the model for $n = 1$ (portion $y \geq 0$ only). The contour interval is $0.1Q/(UL_0)$, where L_0 is defined by equation (2). The heavy black line shows the width predicted by the analytical similarity solution (Methods). **c**, Same as **b**, but for $n = 3.5$.

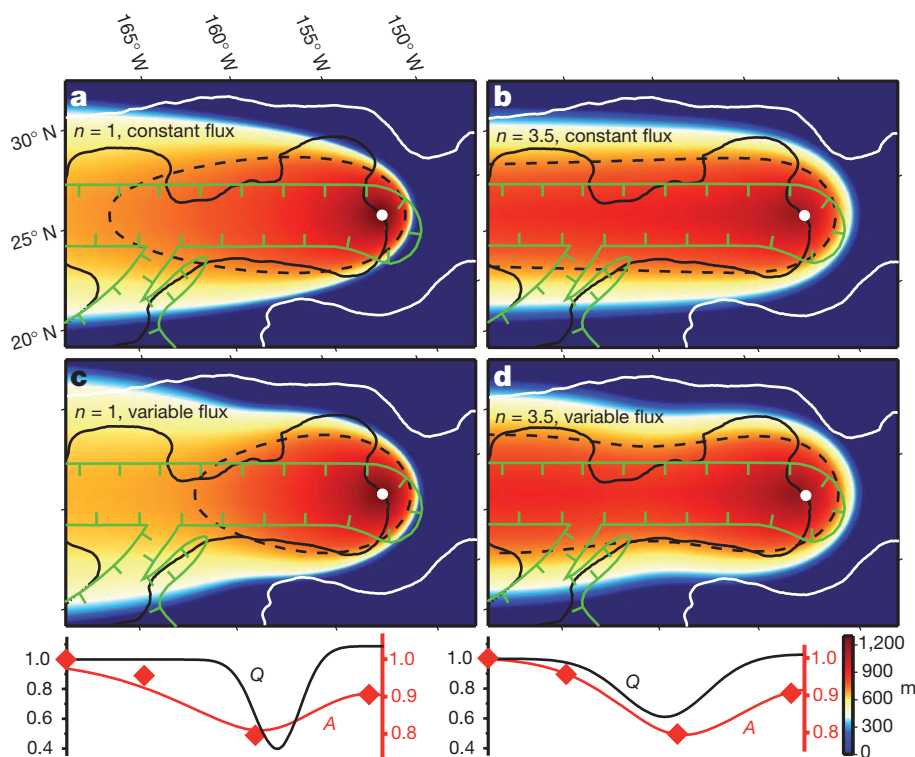


Figure 3 | Comparison of the lubrication model predictions with the residual topography of the Hawaiian swell. Data are shown as oblique Mercator projections. The solid lines are smoothed contours of the residual topography for 0 m (white) and 700 m (black). The green curve bounds the areas excluded from the analysis (Methods). Colours show the solutions of the lubrication equation for diffusion creep ($n = 1$; **a** and **c**), dislocation creep ($n = 3.5$; **b** and **d**), constant plume flux (**a** and **b**) and variable plume flux (**c** and **d**) that best fit the residual topography (Methods). The 700 m contour of each solution is shown by a dotted black line (compare with the solid black line). The flux $Q(t = -x/U)$ and the swell's cross-sectional area $A(x)$ for the two variable-flux cases are also shown below, each normalized to its value at the left of the figure. The diamond symbols are estimates (similarly normalized) of $A(x)$ obtained from the residual topography¹⁸.

low-density material is at depths near or slightly above the base of normal lithosphere. This result, together with the well-defined lateral boundaries of the swell, implies that the region in which our inference of dislocation creep deformation applies is well constrained in all three dimensions. Our result is also consistent with geodynamical modelling of the apparent thermal age of Pacific lithosphere inferred from seismology, which suggests that dislocation creep is the dominant deformation mechanism at depths < 410 km throughout the Pacific upper mantle²³.

We now use the numerical solution of Fig. 3d to estimate the values of the Hawaiian plume buoyancy flux $B \equiv Q\delta\rho$ and the rheological prefactor D in equation (1) (Methods). The best-fitting values of L_0 and S_0 for that solution imply $B = 5,610 \text{ kg s}^{-1}$, within the range (2,800–8,700 kg s^{-1}) of previous estimates^{20,24,25}. The values also imply $D = (8.7\text{--}19.8) \times 10^{-38} \text{ kg m}^{-1} \text{ s}^{-12/7}$, which is more easily understood by translating it into the minimum viscosity η_{\min} within the plume head. For the solution of Fig. 3d and a realistic range of values of $\delta\rho$ (Methods), the maximum strain rate within the plume head is $\dot{\epsilon}_{\max} = (3.1\text{--}3.7) \times 10^{-13} \text{ s}^{-1}$. Equation (1) then implies $\eta_{\min} = (2.3\text{--}3.3) \times 10^{19} \text{ Pa s}$.

The above viscosity estimate can be compared with laboratory-based rheological laws for dry^{26,27} olivine aggregates as a function of temperature, pressure and strain rate³ (Methods). For strain rates $(3.1\text{--}3.7) \times 10^{-13} \text{ s}^{-1}$ and representative values of temperature and pressure in the Hawaiian plume head, the experimental rheological law predicts a viscosity $\eta_{\text{exp}} = (3.2\text{--}15) \times 10^{18} \text{ Pa s}$, a factor 1.6–10.3 lower than η_{\min} .

An alternative mineralogical model for the Hawaiian plume posits that a substantial portion of the source region of Hawaiian lavas is entirely free of olivine, being dominated instead by clinopyroxene²⁸. It is therefore of interest to compare the viscosity inference from our model with the predictions of laboratory deformation experiments on

natural clinopyroxene aggregates⁴, for which $n = 4.7$. The solutions of the lubrication equation with $n = 4.7$ that best fit the Hawaiian residual topography are shown in Supplementary Fig. 3. The maximum strain rate is $\dot{\epsilon}_{\max} = (6.3\text{--}7.6) \times 10^{-13} \text{ s}^{-1}$, which corresponds to a viscosity $\eta_{\min} = (1.0\text{--}1.5) \times 10^{19} \text{ Pa s}$. For the same values of temperature and pressure as previously, the experimental rheological law for clinopyroxene (Methods) predicts a viscosity $\eta_{\text{exp}} = (3.4\text{--}7.5) \times 10^{18} \text{ Pa s}$, a factor of 1.3–4.4 lower than η_{\min} . Our model predictions thus agree slightly better with the laboratory data for clinopyroxene than with those for olivine. However, the error bars are large because the applicability of laboratory-based rheological laws to mantle conditions is difficult to demonstrate. We conclude that our inferred rheology is probably consistent with a source comprising either olivine-rich or clinopyroxene-rich components, or both.

METHODS SUMMARY

The lubrication model with constant σ was validated against a three-dimensional numerical code for convection in a fluid with viscosity dependent on temperature, pressure and strain rate (Supplementary Figs 1 and 2). The lubrication equation was solved numerically using an explicit finite-difference method. The scales L_0 and S_0 that yield the best fit of a given dimensionless solution of the lubrication equation to the residual topography data were determined by maximizing the variance reduction. For non-steady-state solutions, the time-varying flux $Q(t)$ was also adjusted iteratively to obtain the best possible fit between the cross-sectional area distribution $A(x)$ of the swell predicted by the lubrication solution and that estimated from the observations.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 September 2010; accepted 7 March 2011.

Published online 11 May 2011.

1. Poirier, J.-P. *Creep of Crystals* (Cambridge University Press, 1985).
2. Karato, S.-I. *Deformation of Earth Materials* (Cambridge University Press, 2008).

3. Keefner, J. W., Mackwell, S. J., Kohlstedt, D. L. & Heidelbach, F. Dependence of the creep of dunite on oxygen fugacity: implications for viscosity variations in Earth's mantle. *J. Geophys. Res.* doi:10.1029/2010JB00748 (in the press).
4. Bystricky, M. & Mackwell, S. Creep of dry clinopyroxene aggregates. *J. Geophys. Res.* **106**, 13443–13454 (2001).
5. Bai, Q., Mackwell, S. & Kohlstedt, D. L. High-temperature creep of olivine single crystals. 1. Mechanical results for buffered samples. *J. Geophys. Res.* **96**, 2441–2463 (1991).
6. Peltier, W. R. Glacial isostatic adjustment. 2: Inverse problem. *Geophys. J. R. Astron. Soc.* **46**, 669–705 (1976).
7. Hager, B. H., Clayton, R. W., Richards, M. A., Comer, R. P. & Dziewonski, A. M. Lower mantle heterogeneity, dynamic topography and the geoid. *Nature* **313**, 541–545 (1985).
8. Mitrova, J. X. & Forte, A. M. A new inference of mantle viscosity based upon joint inversion of convection and glacial isostatic adjustment data. *Earth Planet. Sci. Lett.* **225**, 177–189 (2004).
9. Karato, S., Jung, H., Katayama, I. & Skemer, P. Geodynamic significance of seismic anisotropy of the upper mantle: new insights from laboratory studies. *Annu. Rev. Earth Planet. Sci.* **36**, 59–95 (2008).
10. Montagner, J. P. & Tanimoto, T. Global upper mantle tomography of seismic velocities and anisotropies. *J. Geophys. Res.* **96**, 20337–20351 (1991).
11. Long, M. D. & Silver, P. G. Shear wave splitting and mantle anisotropy: measurements, interpretations, and new directions. *Surv. Geophys.* **30**, 407–461 (2009).
12. Wilson, J. T. A possible origin of the Hawaiian islands. *Can. J. Phys.* **41**, 863–870 (1963).
13. Morgan, W. J. Convection plumes in the lower mantle. *Nature* **230**, 42–43 (1971).
14. Montelli, R. *et al.* Finite-frequency tomography reveals a variety of plumes in the mantle. *Science* **303**, 338–343 (2004).
15. Wolfe, C. J. *et al.* Mantle shear-wave velocity structure beneath the Hawaiian hot spot. *Science* **326**, 1388–1390 (2009).
16. Watts, A. B. & Cochran, J. R. Gravity anomalies and flexure of the lithosphere along the Hawaiian-Emperor Seamount Chain. *Geophys. J. R. Astron. Soc.* **38**, 119–141 (1974).
17. Tarduno, J. A. *et al.* The Emperor seamounts: southward motion of the Hawaiian hotspot plume in Earth's mantle. *Science* **301**, 1064–1069 (2003).
18. Davies, G. F. Temporal variation of the Hawaiian plume flux. *Earth Planet. Sci. Lett.* **113**, 277–286 (1992).
19. Olson, P. in *Magma Transport and Storage* (ed. Ryan, M.) 33–51 (John Wiley, 1990).
20. Ribe, N. M. & Christensen, U. Three-dimensional modelling of plume-lithosphere interaction. *J. Geophys. Res.* **99**, 669–682 (1994).
21. Marks, K. M. & Sandwell, D. T. Analysis of geoid height versus topography for oceanic plateaus and swells using nonbiased linear regression. *J. Geophys. Res.* **96**, 8045–8055 (1991).
22. Cserepes, L., Christensen, U. & Ribe, N. M. Geoid height versus topography for a plume model of the Hawaiian swell. *Earth Planet. Sci. Lett.* **178**, 29–38 (2000).
23. van Hunen, J., Zhong, S., Shapiro, N. M. & Ritzwoller, M. H. New evidence for dislocation creep from 3-D geodynamic modeling of the Pacific upper mantle structure. *Earth Planet. Sci. Lett.* **238**, 146–155 (2005).
24. Sleep, N. H. Hotspots and mantle plumes: some phenomenology. *J. Geophys. Res.* **95**, 6715–6736 (1990).
25. Ribe, N. M. & Christensen, U. The dynamical origin of Hawaiian volcanism. *Earth Planet. Sci. Lett.* **171**, 517–531 (1999).
26. Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle: implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144**, 93–108 (1996).
27. Karato, S.-I. Insights into the nature of plume-asthenosphere interaction from central Pacific geophysical anomalies. *Earth Planet. Sci. Lett.* **274**, 234–240 (2008).
28. Sobolev, A. V., Hofmann, A. W., Sobolev, S. V. & Nikogosian, I. K. An olivine-free mantle source of Hawaiian shield basalts. *Nature* **434**, 590–597 (2005).
29. Smith, W. H. F. & Sandwell, D. T. Global seafloor topography from satellite altimetry and ship depth soundings. *Science* **277**, 1956–1962 (1997).
30. Stein, C. & Stein, S. A model for the global variation in oceanic depth and heat flow with lithospheric age. *Nature* **359**, 123–129 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Davaille, C. Herzberg, S.-I. Karato and D. Kohlstedt for discussions and advice. This work was supported by the French embassy in Tehran and by the SEDIT programme of INSU and the ANR (grant PTECTO) in France.

Author Contributions N.A. derived the lubrication equation, determined the similarity solution and the full numerical solutions of that equation, and analysed the topography data. N.M.R. proposed the idea for the study, determined the three-dimensional numerical solutions with temperature-dependent rheology, and wrote the manuscript. F.S. co-directed the parts of the work done in Zanjan. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.A. (n_asaadi@iasbs.ac.ir) or N.M.R. (ribe@fast.u-psud.fr).

METHODS

Derivation and solution of the lubrication equation. In the following, the subscripts α and β take on the values 1 or 2, corresponding to vector components in the x - and y -directions, respectively.

Because the plume head is thin, the pressure inside it is nearly hydrostatic. Its value relative to the pressure outside is:

$$p = g\delta\rho(h - z) \quad (5)$$

where g is the gravitational acceleration and $-\delta\rho$ is the density deficit of the plume relative to the ambient mantle. The horizontal flow in the plume head is a Poiseuille flow driven by the lateral gradient of the pressure. In the lubrication approximation, the horizontal velocity components u_α satisfy:

$$\partial_z(\eta\partial_z u_\alpha) = \partial_\alpha p \equiv g\delta\rho\partial_z h \quad (6)$$

where the viscosity η is given by equation (1). The solution of equation (6) for which the slip $u_\alpha - U\delta_{1\alpha}$ at the plume head's upper surface $z = 0$ and the shear traction $\eta\partial_z u_\alpha$ at its lower surface $z = h$ both vanish is:

$$u_\alpha = U\delta_{1\alpha} - \frac{D}{n+1}(g\delta\rho)^n |\nabla h|^{n-1} [h^{n+1} - (h-z)^{n+1}] \partial_\alpha h \quad (7)$$

where $\delta_{1\alpha}$ is the Kronecker delta symbol.

Conservation of mass over the whole thickness of the plume head requires:

$$\partial_t h + \partial_\beta \int_0^h u_\beta dz = \frac{Q}{\pi a^2} \exp(-r^2/a^2) \quad (8)$$

where the source of buoyant fluid has been represented by a radially symmetric Gaussian distribution of width a with $r^2 = x^2 + y^2$, and summation over the repeated subscript β is understood. Substitution of equation (7) into equation (8) finally yields:

$$\partial_t h + U\partial_x h = \nabla \cdot [\sigma |\nabla h|^{n-1} \nabla (h^{n+3})] + \frac{Q}{\pi a^2} \exp(-r^2/a^2) \quad (9)$$

where $\sigma = D(g\delta\rho)^n / [(n+2)(n+3)]$.

Equation (9) was prepared for numerical solution by rewriting it in terms of the dimensionless variables hUL_0/Q , x/L_0 , y/L_0 and tUL_0 . Solutions were obtained using an explicit finite-difference method. The gravitational spreading term (first term on the right-hand side) was discretized using a centred difference scheme with fourth-order accuracy. The advection term $U\partial_x h$ was discretized using second-order accurate upwind differencing. Time stepping was performed explicitly using a second-order accurate midpoint method.

Far downstream from the hotspot, the source term in equation (9) is negligible, and gravitational spreading is primarily in the y direction. The steady-state form of equation (9) then takes the form:

$$U\partial_x h = \sigma(n+3)\partial_y (h^{n+2} |\partial_y h|^{n-1} \partial_y h) \quad (10)$$

where σ has been assumed constant. Conservation of the downstream volume flux requires:

$$U \int_{-L}^L h dy = Q \quad (11)$$

where $L(x)$ is the half-width of the plume head. Equations (10) and (11) admit a similarity solution of the form:

$$h = \frac{Q}{UL(x)} H(\zeta) \quad (12)$$

where $\zeta = \frac{y}{L(x)}$. The appropriate boundary conditions are:

$$0 = L(x_0) = H'(0) = H(1) \quad (13)$$

where x_0 is the virtual position of the hotspot as seen by the solution far downstream and the prime indicates $d/d\zeta$. Substitution of equation (12) into equations (10) and (11) yields ordinary differential equations for $L(x)$ and $H(\zeta)$ that can be solved subject to the conditions (13) to yield:

$$H(\zeta) = c_1 \left(1 - |\zeta|^{\frac{1+n}{2n+1}} \right)^{\frac{n}{2n+1}} \quad (14)$$

and

$$L = c_2 L_0 \left(\frac{x - x_0}{L_0} \right)^{\frac{1}{3n+2}} \quad (15)$$

where

$$c_1 = \frac{\Gamma\left(\frac{5n^2+5n+1}{(n+1)(2n+1)}\right)}{2\Gamma\left(\frac{2n+1}{n+1}\right)\Gamma\left(\frac{3n+1}{2n+1}\right)} \quad (16)$$

$$c_2 = \left[(n+3)(3n+2) \left(\frac{n+1}{2n+1} \right)^n c_1^{2n+1} \right]^{\frac{1}{3n+2}} \quad (17)$$

and Γ is the Gamma function. For $n = 3.5$, $c_1 \approx 0.6775$ and $c_2 \approx 0.9432$. The virtual hotspot position x_0 is determined by least-squares fitting of equation (15) to the full numerical solution of equation (9), yielding $x_0/L_0 = 0.753$ for $n = 1$ and 0.272 for $n = 3.5$.

Validation of the lubrication model. Although the gravitational spreading parameter σ depends on temperature and pressure in general, we treat it as constant in this study. Here we justify this assumption using a three-dimensional numerical code^{20,31} for convection in a fluid whose viscosity η depends on temperature, pressure and strain rate according to:

$$\eta = \eta_{\min} + (\eta_{\max}^{-1} + \eta_1^{-1})^{-1}, \quad (18)$$

$$\eta_1 = \eta_0 \left(\frac{\dot{\epsilon}_1}{\dot{\epsilon}_0} \right)^{\frac{1-n}{n}} \exp\left(\frac{E+pV}{nRT}\right), \quad (19)$$

$$\dot{\epsilon}_1 = (I^2 + I_{\min}^2)^{1/2} \quad (20)$$

where $I = (2e_{ij}e_{ij})^{1/2}$ is the second invariant of the strain rate tensor e_{ij} . In equations (18) to (20), $\eta_0 = 10^{21}$ Pa s is a reference viscosity, and $\eta_{\min} = 0.001\eta_0$ and $\eta_{\max} = 500\eta_0$ are the minimum and maximum allowable viscosities, respectively. $\dot{\epsilon}_0 = 2 \times 10^{-15}$ s⁻¹ is a reference strain rate, and $I_{\min} = 10^{-15}$ s⁻¹ is the minimum allowable value of I . Finally, T is the absolute temperature, p is the pressure, E is the activation energy, V is the activation volume, and R is the universal gas constant.

The equations of conservation of mass, momentum and energy are solved using a hybrid spectral (in the two horizontal directions) and finite-difference (in the vertical direction) method in which the coupling of different spectral modes by lateral viscosity variations is calculated iteratively^{20,31}. A thermal plume is generated by imposing a hot temperature anomaly $\Delta T_{\max} \exp(-r^2/a^2)$ on the bottom of the model box (depth $d = 400$ km). The thermal plume interacts with the shear flow generated by motion of the upper surface of the model box at a constant velocity $U = 2.7 \times 10^{-9}$ m s⁻¹ (8.6 cm yr⁻¹). Effects of melting-induced depletion²⁵ are neglected for simplicity.

Supplementary Fig. 1a shows the steady-state temperature field for a plume with buoyancy flux $B = 2,340$ kg s⁻¹ and $\Delta T_{\max} = 200$ K in a fluid with diffusion creep ($n = 1$) rheology with $E = 400$ kJ mol⁻¹ and $V = 6.1 \times 10^{-6}$ m³ mol⁻¹. To compare this three-dimensional solution quantitatively with the lubrication model, we use the former to calculate the isostatic topography:

$$S_{\text{iso}} = \frac{\rho_0 \alpha}{\rho_0 - \rho_w} \int_0^d \delta T dz \quad (21)$$

where $\delta T(x, y, z)$ is the temperature anomaly (temperature with the plume present less the temperature in a second solution with $\Delta T_{\max} = 0$) and $\alpha = 3.5 \times 10^{-5}$ K⁻¹ is the coefficient of thermal expansion. Supplementary Fig. 1b shows S_{iso} on the central axis $y = 0$ as a function of $x - x_0$ with $x_0 = 163$ km (black line), together with the power-law decay of equation (4) with $n = 1$ that best fits it (red line). The downstream decay of the topography agrees closely with the prediction $S_{\text{iso}} \propto (x - x_0)^{-0.2}$ of the lubrication model.

Supplementary Fig. 2 is for a plume with the same values of B , ΔT_{\max} , E and V as in Supplementary Fig. 1, but with a dislocation creep rheology ($n = 3.5$). Again, the topography decays downstream in accordance with the lubrication prediction $S_{\text{iso}} \propto (x - x_0)^{-0.08}$.

Supplementary Figs 1 and 2 show that the lubrication model accurately reproduces the dynamics of the more complicated three-dimensional code for both diffusion creep ($n = 1$) and dislocation creep ($n = 3.5$) rheologies, even in the presence of strongly temperature-dependent viscosity. The physical reason for this is that the largest temperature variations in the three-dimensional code are confined near the upwelling plume conduit, so that lateral variations of temperature throughout most of the plume head are relatively small (Supplementary Figs 1a and 2a). A lubrication model with constant σ can therefore be used instead of the full three-dimensional code, which, because of its high computational cost and multiplicity of parameters, would render intractable the task of finding the model with the optimal fit to the residual topography data.

Comparison of the model predictions with the residual topography. The first step is to exclude those portions of the residual topography that are unrelated to the swell, which include the Hawaiian islands, the surrounding moat produced by lithospheric flexure, and the nearby Mid-Pacific mountains. These areas are enclosed by the green lines in Fig. 3. We also excluded all topography outside the range 0–1,300 m.

Once a numerical solution of equation (9) was found for a given value of n , we used a least-squares procedure (maximization of variance reduction) to determine the values L_0^{bf} and S_0^{bf} of the length scale L_0 and the topography scale S_0 such that the re-dimensionalized solution best fits the non-excluded residual topography. For a steady-state solution with constant Q , the determination is direct. For non-steady-state solutions in which Q is a function of time, we determined L_0^{bf} and S_0^{bf} iteratively by adjusting the shape of the function $Q(t)$ to obtain the best possible agreement between the cross-sectional area distribution $A(x)$ of the swell topography predicted by the numerical model and that estimated directly from the residual topography data¹⁸. An iterative procedure is required because the volume flux $Q(t)$ of the Hawaiian plume varies on a timescale that is comparable to the fluid-mechanical adjustment time L_0/U , which implies that the cross-sectional area of the swell at a given distance x from the hotspot is not simply proportional to the plume flux at the earlier time $t = -x/U$ when the cross-section was directly over the hotspot. For the non-steady lubrication solution of Fig. 3d, $L_0^{\text{bf}} = 630$ km and $S_0^{\text{bf}} = 1,375$ m.

Given L_0^{bf} and S_0^{bf} , the buoyancy flux B and the rheological prefactor D of the best-fitting lubrication solution were found by solving the two simultaneous equations $L_0 = L_0^{\text{bf}}$ and $S_0 = S_0^{\text{bf}}$, where L_0 and S_0 are defined by equations (2) and (3) respectively. The results are:

$$B = S_0^{\text{bf}} L_0^{\text{bf}} (\rho_0 - \rho_w) U \quad (22)$$

$$D = \frac{(2+n)(3+n) \left(L_0^{\text{bf}} / g \right)^n \delta \rho^{1+n} U}{[S_0^{\text{bf}} (\rho_0 - \rho_w)]^{1+2n}} \quad (23)$$

To evaluate equations (22) and (23) we assumed $U = 2.7 \times 10^{-9}$ m s⁻¹ and $\delta \rho = \rho_0 \alpha \Delta T$, where $\alpha = 3.5 \times 10^{-5}$ K⁻¹ and $\Delta T = 125$ –150 K is a typical excess

temperature in the plume head far from the hotspot, based on a maximum temperature contrast (250–300 K) of the upwelling material beneath the hotspot³².

Comparison of the model predictions with laboratory experiments. Laboratory deformation experiments in the dislocation creep regime typically yield rheological laws of the form:

$$\dot{\epsilon} = \mathcal{A} f^m \tau^n \exp\left(-\frac{E+pV}{RT}\right) \quad (24)$$

where \mathcal{A} is a constant and f is the oxygen fugacity. To evaluate equation (24) within the Hawaiian plume head downstream from the hotspot, we assume $p = 3.3$ GPa (corresponding to a depth ≈ 100 km) and $T = 1,748$ –1,773 K, corresponding to an excess temperature $\Delta T = 125$ –150 K relative to an ambient mantle temperature 1,350°C. For dry olivine aggregates, $m = 0.2$, $n = 3.59$, $\mathcal{A} = 1.15 \times 10^{-19}$ s⁻¹ Pa^{-n-m}, $E = 449$ kJ mol⁻¹ (ref. 3), and $V = 1.7$ – 2.8×10^{-5} m³ mol⁻¹ (refs 33 and 34). In addition, we assume $f \in [10^{-6.2}, 10^{-2.1}]$ Pa (ref. 3). For dry natural clinopyroxene aggregates, $m = 0$, $n = 4.7$, $\mathcal{A} = 4.0 \times 10^{19}$ s⁻¹ Pa⁻ⁿ, and $E + pV = 760$ kJ mol⁻¹ for an average confining pressure $P = 425$ MPa (ref. 4). Because V is poorly known for clinopyroxene, we assume the same range of values as for olivine.

31. Christensen, U. & Harder, H. 3-D convection with variable viscosity. *Geophys. J. Int.* **104**, 213–226 (1991).
32. Herzberg, C. & Asimow, P. D. Petrology of some oceanic island basalts: PRIMELT2.XLS software for primary magma calculation. *Geochem. Geophys. Geosyst.* **9**, Q09001 (2008).
33. Kirby, S. H. Rheology of the lithosphere. *Rev. Geophys.* **21**, 1458–1487 (1983).
34. Borch, R. S. & Green, H. W. II. Deformation of peridotite at high pressure in a new molten salt cell: comparison of traditional and homologous temperature treatments. *Phys. Earth Planet. Inter.* **55**, 269–276 (1989).

Profound early control of highly pathogenic SIV by an effector memory T-cell vaccine

Scott G. Hansen¹, Julia C. Ford¹, Matthew S. Lewis¹, Abigail B. Ventura¹, Colette M. Hughes¹, Lia Coyne-Johnson¹, Nathan Whizin¹, Kelli Oswald², Rebecca Shoemaker², Tonya Swanson¹, Alfred W. Legasse¹, Maria J. Chiuchio³, Christopher L. Parks³, Michael K. Axthelm¹, Jay A. Nelson¹, Michael A. Jarvis¹, Michael Piatak Jr², Jeffrey D. Lifson² & Louis J. Picker¹

The acquired immunodeficiency syndrome (AIDS)-causing lentiviruses human immunodeficiency virus (HIV) and simian immunodeficiency virus (SIV) effectively evade host immunity and, once established, infections with these viruses are only rarely controlled by immunological mechanisms^{1–3}. However, the initial establishment of infection in the first few days after mucosal exposure, before viral dissemination and massive replication, may be more vulnerable to immune control⁴. Here we report that SIV vaccines that include rhesus cytomegalovirus (RhCMV) vectors⁵ establish indefinitely persistent, high-frequency, SIV-specific effector memory T-cell (T_{EM}) responses at potential sites of SIV replication in rhesus macaques and stringently control highly pathogenic SIV_{MAC239} infection early after mucosal challenge. Thirteen of twenty-four rhesus macaques receiving either RhCMV vectors alone or RhCMV vectors followed by adenovirus 5 (Ad5) vectors (versus 0 of 9 DNA/Ad5-vaccinated rhesus macaques) manifested early complete control of SIV (undetectable plasma virus), and in twelve of these thirteen animals we observed long-term (≥ 1 year) protection. This was characterized by: occasional blips of plasma viraemia that ultimately waned; predominantly undetectable cell-associated viral load in blood and lymph node mononuclear cells; no depletion of effector-site $CD4^+$ memory T cells; no induction or boosting of SIV Env-specific antibodies; and induction and then loss of T-cell responses to an SIV protein (Vif) not included in the RhCMV vectors. Protection correlated with the magnitude of the peak SIV-specific $CD8^+$ T-cell responses in the vaccine phase, and occurred without anamnestic T-cell responses. Remarkably, long-term RhCMV vector-associated SIV control was insensitive to either $CD8^+$ or $CD4^+$ lymphocyte depletion and, at necropsy, cell-associated SIV was only occasionally measurable at the limit of detection with ultrasensitive assays, observations that indicate the possibility of eventual viral clearance. Thus, persistent vectors such as CMV and their associated T_{EM} responses might significantly contribute to an efficacious HIV/AIDS vaccine.

Conventional prime-boost vaccine regimens with non-persistent vectors lead to lymphoid tissue-based memory T-cell responses ('central memory' or T_{CM}), which deliver peak effector responses only after T_{CM} cells have undergone antigen-stimulated expansion, differentiation and trafficking⁶—too late to effectively control pathogens with the rapid replication and spread kinetics and highly developed immune evasion capabilities of the AIDS-causing lentiviruses^{2,4,5}. As T-cell effector responses are likely to be much more effective against the smaller, localized and less diverse viral populations present in the first hours and days of mucosally acquired HIV/SIV infection^{2,4,7,8}, we proposed that a vaccine able to 'pre-position' differentiated effector cells (T_{EM}) at such early replication sites would demonstrate improved efficacy. Such T_{EM} responses are the hallmark of persistent agents^{9,10}, prompting our development of SIV vectors based on the persistent

β -herpesvirus RhCMV. As recently reported⁵ and illustrated in Supplementary Fig. 1, RhCMV/SIV vectors can establish and indefinitely maintain high-frequency SIV-specific, T_{EM} -biased, $CD4^+$ and $CD8^+$ T-cell responses in diverse tissue sites of RhCMV⁺ rhesus macaques, and in a small efficacy study were associated with early control of intrarectally administered SIV_{MAC239}. To evaluate potential differential effects of persistent vector/ T_{EM} -biased versus non-persistent vector/ T_{CM} -biased, SIV-specific T-cell responses on the outcome of mucosal SIV_{MAC239} infection, we compared naturally RhCMV⁺ male rhesus macaques vaccinated with: (1) RhCMV/SIV vectors alone (Group A); (2) RhCMV/SIV vectors followed by replication-defective Ad5 vectors (Group B); and (3) a standard DNA prime/Ad5 vector boost benchmark vaccine (Group C)^{11–13} versus unvaccinated control rhesus macaques (Group D; Fig. 1a). RhCMV/SIV vectors efficiently super-infected all Group A and B macaques and elicited robust $CD4^+$ and $CD8^+$ T-cell responses to all vector-encoded SIV proteins (Fig. 1b and Supplementary Figs 2–4). The Ad5 vector boost of Group B macaques, and the DNA/Ad5 regimen given to Group C macaques were also strongly immunogenic (Fig. 1b and Supplementary Figs 3, 4). Although the pattern of development of the SIV-specific T-cell responses differed between these vectors (Supplementary Fig. 3a), the magnitude of the total SIV-specific, $CD4^+$ and $CD8^+$ T-cell responses at the end of the vaccine phase in Groups A, B and C were similar (Fig. 1b and Supplementary Fig. 4). Consistent with previous results⁵, RhCMV/SIV-vector-elicited, SIV-specific $CD8^+$ T-cell responses exhibited different epitope targeting than the DNA- and/or Ad5-vector-elicited responses (Supplementary Fig. 3b), and maintained a markedly T_{EM} -biased phenotype over the entire vaccine phase, in contrast to the development of a more T_{CM} -biased response in the DNA/Ad5-vaccinated macaques (Supplementary Fig. 5).

At week 59 after initial vaccination, all rhesus macaques were challenged via the intrarectal route with highly pathogenic SIV_{MAC239} using a repeated, limiting dose protocol⁵. The number of challenges required to achieve measureable infection—plasma viral load $>$ threshold ($30 \text{ copies ml}^{-1}$)—was not significantly different between Groups A–D (Supplementary Fig. 6), but the subsequent course of infection in these groups was markedly different (Fig. 1c). Of 28 unvaccinated controls (both concurrent and historical), 27 exhibited typical progressive SIV_{MAC239} infection and one exhibited an initially non-progressive infection (transient viraemia) that spontaneously progressed 105 days later. Similarly, all DNA/Ad5-vaccinated macaques (9 of 9) manifested progressive infection, albeit with reduced mean plasma viral load compared to controls (see later). In contrast, 13 of the 24 rhesus macaques that received RhCMV/SIV vectors (6/12 in Group A; 7/12 in Group B) presented with an initial burst of plasma SIV, ranging in magnitude from as few as 60 to as many as 4×10^7 SIV RNA copies ml^{-1} , which was followed by rapid control to undetectable levels (Fig. 1c, d). From 3–18 weeks after infection, all but one of these

¹Vaccine and Gene Therapy Institute, Departments of Molecular Microbiology and Immunology and Pathology, and the Oregon National Primate Research Center, Oregon Health & Science University, Beaverton, Oregon 97006, USA. ²AIDS and Cancer Virus Program, SAIC Frederick Inc., National Cancer Institute-Frederick, Frederick, Maryland 21702, USA. ³International AIDS Vaccine Initiative, Vaccine Design and Development Laboratory, 140 58th Street, Building A, Unit 8J, Brooklyn, New York 11220, USA.

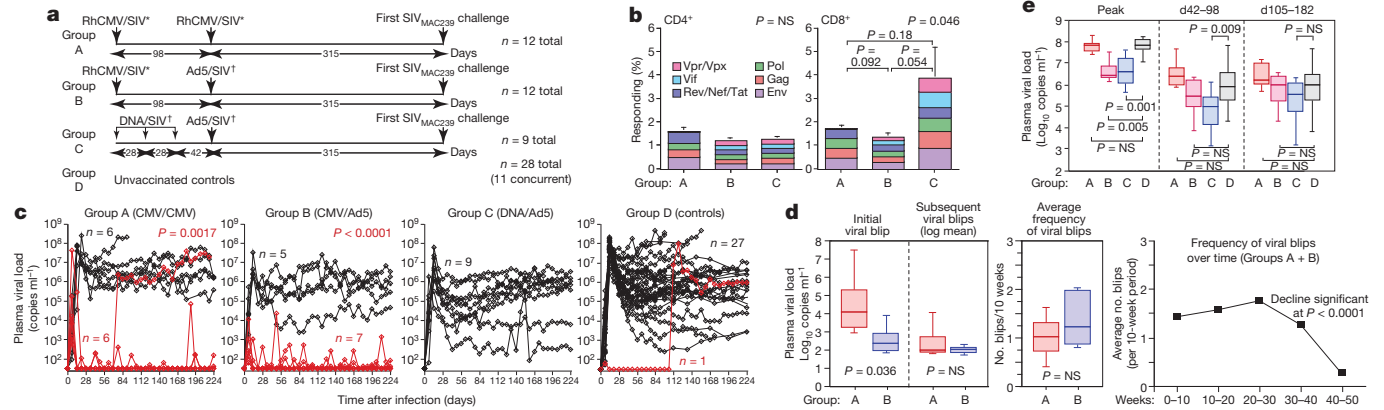


Figure 1 | Immunogenicity and efficacy of RhCMV/SIV vectors.

a, Schematic of the vaccination protocols used in this study. Note that all rhesus macaques used were naturally RhCMV⁺ at the start of the study. Asterisk indicates Gag + Rev/Nef/Tat + Env + Pol. Dagger indicates pan-SIV proteome. **b**, Comparison of the mean frequency (± s.e.m.) of the overall SIV-specific CD4⁺ and CD8⁺ T-cell responses and the contribution of the designated SIV proteins to these total responses in the blood memory compartments of Groups A–C rhesus macaques at the end of the vaccine phase. The Kruskal–Wallis test was used to determine the significance of differences in total SIV-specific response frequencies among the three vaccine groups, with the Wilcoxon rank sum test used to perform pair-wise analysis for the CD8⁺ response. As these latter *P* values were >0.05, we concluded that overall response frequencies of Groups A, B and C were not significantly different. NS, not significant. *P* values in top right corners of graphs are for overall Group A versus B versus C. **c**, Outcome of repeated, limiting dose, intrarectal SIV_{MAC239}

protected rhesus macaques demonstrated one or more repeat episodes of transient viraemia that were always controlled back down to below detection limits (Fig. 1c, d). These periodic viral blips were similar in magnitude in Group A and B controllers, and recurred on average about once every 7 weeks during the first 30 weeks after infection (Fig. 1d). Notably, the frequency of these viral blips declined significantly after week 30 such that by 52 weeks after infection, viral blips were rarely observed (Fig. 1d). No SIV-mediated pathogenesis (loss of effector site CD4⁺ T cells) was noted in Group A and B controllers (Supplementary Fig. 7), and the vast majority of blood and lymph node mononuclear cell specimens from these macaques were negative for cell-associated SIV RNA and DNA (Supplementary Fig. 8). Six of 12 Group A and 5 of 12 Group B rhesus macaques were not protected in this novel manner, but rather, demonstrated a typical pattern of progressive infection with associated pathogenesis (Fig. 1c, e and Supplementary Fig. 7). The mean peak and plateau phase plasma viral loads of the Group A rhesus macaques with progressive infection were not statistically different from Group D controls (Fig. 1e), indicating that once systemic, progressive infection was established, RhCMV/SIV-vector-elicited responses were unable to control virus replication. The addition of Ad5/SIV vectors in the Group B vaccination regimen was associated with a significantly reduced peak viraemia in Group B macaques with progressive infection compared to Group D controls, but this difference was lost in plateau phase. Consistent with previous reports^{11–13}, the benchmark DNA/Ad5-vaccinated macaques (Group C) showed significantly reduced log mean peak and early plateau phase (6–14 weeks after infection) plasma viral loads, but for most of these macaques this partial virological control was short-lived, as log mean plasma viral loads in later plateau phase were also not different from Group D controls (Fig. 1c, e). Importantly, the stringent control of SIV infection in protected Group A and B rhesus macaques was not associated with CD8⁺ T-cell responses restricted by protective MHC alleles (Supplementary Fig. 3b) or with TRIM5 polymorphisms associated with target cell susceptibility to SIV infection (Supplementary Fig. 9).

Taken together, these data indicate that RhCMV/SIV-vector-elicited immune responses mediate a novel pattern of protection in

challenge of Groups A–D. The significance of differences in the fraction of infected rhesus macaques in each group that met controller criteria (see Methods) was determined by Fisher's exact test (closed symbols in Group D are concurrent controls; open, previous controls given the same challenge; red, controllers; black, non-controllers). **d**, Analysis of the magnitude and frequency of plasma viral load 'blips' in Group A and B controllers over the first 50 weeks of infection, with the significance of the differences in blip magnitude and frequency between Groups A and B determined by the Wilcoxon rank sum test, and the significance of the decline in blip frequency of Group A + B macaques after 30 weeks post-infection determined by analysis of variance and linear trend tests. **e**, Comparison of plasma viral loads in Groups A–D rhesus macaques with progressive infection (excluding Group A and B controllers and Group D macaques with protective MHC alleles not represented in Groups A–C) with the significance of differences between Groups A, B and C versus Group D determined by the Wilcoxon rank sum test.

which mucosally administered SIV_{MAC239} is stringently controlled before the onset of progressive, systemic infection. As shown in Fig. 2a and Supplementary Fig. 10, the peak frequencies of SIV-specific CD8⁺ (but not CD4⁺) T cells during the vaccine phase (which occurred shortly after the boost), but not the frequencies immediately pre-challenge, significantly correlated with protection in both Groups A and B. These peak responses reflect the level of overall production of SIV-specific CD8⁺ T cells by the vaccine, and for a T_{EM}-biased response would probably parallel the extent of T_{EM} seeding at effector sites. SIV Env-specific antibody responses are not generated by our RhCMV/SIV vectors⁵, and did not develop after SIV infection in Group A controllers (Fig. 2b). Although Ad5/SIV-Env-vector-vaccinated rhesus macaques in Group B developed low-titre SIV Env-specific (tissue-culture-adapted SIV_{MAC251}-neutralizing) antibody responses before challenge, these titres did not predict control and were not boosted by controlled infection. In contrast, with the exception of rapid progressors, SIV Env-specific antibody responses developed or were boosted in all macaques with systemic, progressive SIV infection. These findings indicate that antibody responses are unlikely to significantly contribute to the protection observed in Group A and B macaques, and further confirm the stringency of protection in RhCMV/SIV-vector-vaccinated controllers, as SIV replication in these macaques produced insufficient antigen to drive humoral immune responses.

We next investigated the effect of SIV infection on the magnitude of the vaccine-elicited T-cell responses. Notably, Group A rhesus macaques showed an almost complete lack of anamnestic SIV Gag-specific CD4⁺ or CD8⁺ T-cell response to either progressive or controlled SIV infection (Fig. 2c and Supplementary Fig. 11). Group B macaques demonstrated a modest anamnestic response in the setting of control, whereas in the setting of progressive infection they manifested a robust anamnestic response, similar to or only slightly less than that observed in Group C macaques. Thus, despite the facts that Group B macaques manifested circulating SIV-specific CD8⁺ T-cell responses with the characteristic marked T_{EM} bias of RhCMV/SIV-vector-elicited responses (Supplementary Fig. 5), and the early, abrupt RhCMV/SIV-vector-associated pattern of protection (Fig. 1c), these

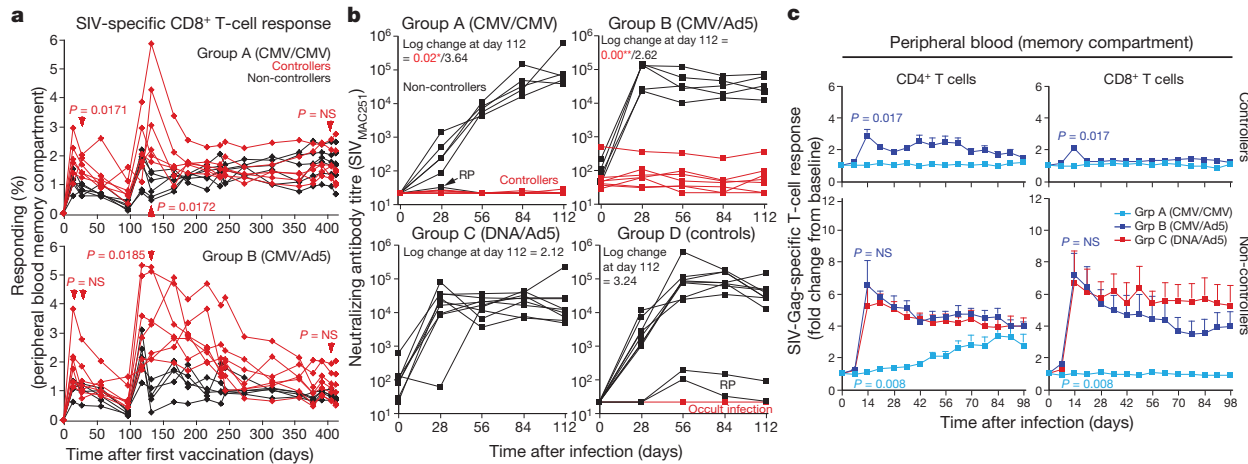


Figure 2 | Immunological correlates of RhCMV/SIV-vector-associated control. **a**, Analysis of total SIV-specific CD8⁺ T-cell responses (SIV Gag + Rev/Nef/Tat + Pol + Env) in the blood memory compartment during the vaccine phase of Group A and B rhesus macaques with differences in the magnitude of these responses between controllers and non-controllers at the designated time points determined by the Wilcoxon rank sum test. **b**, Comparison of the anti-SIV-Env antibody titres in plasma (as measured by neutralization of tissue-culture-adapted SIV_{MAC251}) before and after infection of controller versus non-controller rhesus macaques among Groups A–C and the concurrent Group D macaques. RP, rapid progressor. Occult infection

refers to the initially non-progressive infection (Fig. 1c). The significance of the differences in log change in antibody titre from pre-infection to day 112 post-infection in Group A and B controllers versus Group C macaques was determined by the Wilcoxon rank sum test. * $P < 0.0001$, ** $P < 0.005$. **c**, Analysis of the change in the SIV-Gag-specific CD4⁺ and CD8⁺ T-cell response frequency after controlled versus progressive infection in Groups A, B and C with the significance of differences in peak response boosting between the designated groups determined by the Wilcoxon rank sum test. Error bars show mean \pm s.e.m. P values are compared to Group A for the controllers (top panels) and to Group C for the non-controllers (bottom panels).

macaques seemed to maintain a distinct Ad5-vector-elicited, SIV-specific T_{CM} population capable of anamnestic expansion upon either controlled or progressive SIV infection. Importantly, Group A and B controllers robustly responded to infection with *de novo* (Group A) or boosted (Group B) CD4⁺ and CD8⁺ T-cell responses to SIV Vif, an antigen not included in the RhCMV/SIV vectors used in this study (Supplementary Fig. 12), confirming both the presence of SIV infection in these macaques, and the normal ability of their naive T-cell (Group A) and T_{CM} (Group B) compartments to respond to the infection. These results indicate that not only does RhCMV/SIV-vector-associated viral control occur in the absence of an overt anamnestic response, but that the SIV-specific T_{EM} populations generated by RhCMV/SIV vectors alone seem unable to significantly expand after infection, regardless of whether antigen levels are limiting (controlled infection) or abundant (progressive infection). This lack of anamnestic expansion may account for the inability of Group A macaques (in contrast to Group B macaques) to manifest any suppression of viral replication once a systemic, progressive infection was established.

The decline in the frequency of SIV RNA blips in the plasma of RhCMV/SIV-vector-vaccinated controllers over time suggests progressive loss of SIV-infected cells, either by immune clearance, virolysis or other attritive mechanisms. To explore the extent of residual infection in long-term RhCMV/SIV-vaccinated controllers, we used monoclonal antibodies to deplete CD4⁺ or CD8⁺ lymphocytes from two Group A and two Group B controllers, in comparison to a Group C (DNA/Ad5) and a Group D (unvaccinated) macaque with partial virological control, for each treatment. Administration of the anti-CD4 huOKT4A monoclonal antibody depleted CD4⁺ T cells, but did not increase plasma viraemia in either Group C and D partial controllers or Group A and B complete controllers (Supplementary Fig. 13). In keeping with previous studies^{14,15}, CD8⁺ lymphocyte depletion with the cM-T807 monoclonal antibody did result in a pronounced increase in plasma viral load in Group C and D rhesus macaques with partial control, associated with a robust expansion of SIV Vif-specific CD4⁺ T cells in effector sites (Fig. 3a). In contrast, CD8⁺ lymphocyte depletion failed to increase plasma viraemia in RhCMV/SIV-vector-vaccinated controllers,

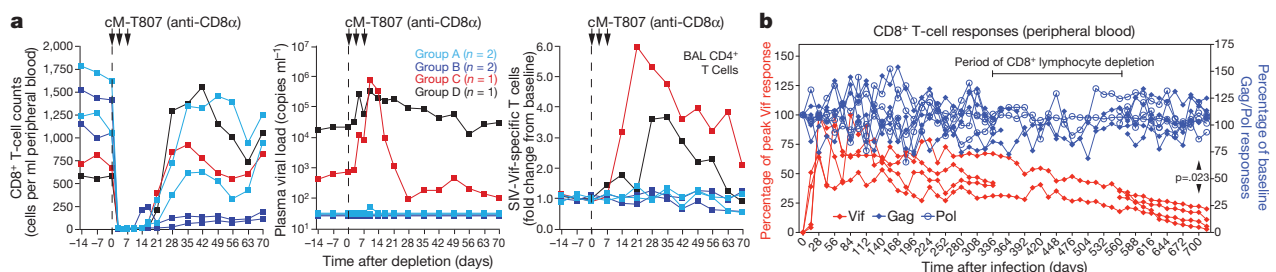


Figure 3 | Immunological characterization of long-term control associated with RhCMV/SIV vector vaccination. **a**, Analysis of the effect of depletion of CD8⁺ lymphocytes with cM-T807 monoclonal antibody on viral replication and boosting of SIV Vif-specific T-cell responses (in the non-depleted CD4⁺ subset) in four long-term RhCMV/SIV-vector-vaccinated controllers (two Group A and two Group B rhesus macaques) versus two conventional controllers (one Group C, DNA/Ad5-vaccinated controller; one Group D spontaneous controller). BAL, broncho-alveolar lavage lymphocytes. **b**, Analysis of the frequencies of blood CD8⁺ T cells specific for SIV proteins that were (Gag, Pol) or were not (Vif) included in the CMV/SIV vectors in the four Group A controllers for which long-term data are available. The response frequencies were normalized to the

response frequencies immediately before SIV infection for the Gag- and Pol-specific responses, and to the peak frequencies following SIV infection for the Vif-specific responses. The four rhesus macaques used in this long-term response analysis include those subjected to transient CD4⁺ or CD8⁺ lymphocyte depletion (two each). As antigen-specific CD8⁺ T-cell responses cannot be reliably determined during the period of overall CD8⁺ lymphocyte depletion, these periods are shown as gaps for two affected rhesus macaques. The significance of differences in the maintenance of response frequencies of Gag- and Pol- versus Vif-specific CD8⁺ T cells in these rhesus macaques was determined by Wilcoxon rank sum analysis.

and the SIV Vif-specific CD4⁺ T-cell responses in these macaques were unchanged after depletion, suggesting the absence of even a transient increase in viral replication not detectable by plasma viral load measurements. These studies extend our previous data on the insensitivity of RhCMV/SIV-vector-associated control to CD8⁺ lymphocyte depletion⁵ to rhesus macaques that manifested a higher initial viraemia as well as a period of subsequent, intermittent plasma viral load blips.

As CD8⁺ T-cell depletion with cM-T807 monoclonal antibody is typically not complete in tissues (Supplementary Fig. 14), lack of viral rebound after such treatment of RhCMV/SIV-vector-vaccinated controllers may simply reflect the potent antiviral function of such residual SIV-specific CD8⁺ T_{EM} cells or, possibly, the compensatory activity of antiviral CD4⁺ T_{EM} cells. On the other hand, these observations also raise the possibility that the frequency of SIV-infected and potentially infectious cells in long-term RhCMV/SIV-vector-vaccinated controllers might have been reduced over time to levels that made detectable viral rebound unlikely. In this regard, we found that in Group A controllers, both CD8⁺ and CD4⁺ T-cell responses to SIV Vif, an antigen that was not included in the RhCMV/SIV vectors and therefore only available from SIV-infected cells, progressively waned over time to an average of <10% of their peak response immediately after (controlled) infection (Fig. 3b and Supplementary Fig. 15). This observation indicates that the numbers of productively infected cells present in these long-term controller macaques are very few, below the threshold necessary to support the initially high-frequency Vif-specific responses. To further examine the extent of residual infection in long-term RhCMV/SIV-vaccinated controllers, we rigorously quantified SIV RNA and DNA at necropsy in four such macaques (\geq week 52 after infection; lacking plasma viral load blips for \geq 10 weeks before necropsy) in comparison to an uninfected macaque, two macaques with SIV infections that were well controlled by standard criteria, and an additional macaque with poorly controlled, progressive SIV infection. As shown in Fig. 4, extensive analysis of lymphoid tissues and immune effector sites of the

RhCMV/SIV-vector-vaccinated controllers with ultra-sensitive nested, quantitative reverse transcription polymerase chain reaction (RT-PCR) and polymerase chain reaction (PCR) assays (10 reactions per tissue specimen) demonstrated that cell-associated SIV RNA and DNA were undetectable (0/10 reactions positive) in 72% and 80% of specimens, respectively. In those tissues where viral sequences were detected, the levels were extremely low (approximately one copy per 10⁷–10⁸ cell equivalents). Notably, the majority of specimens with detectable SIV DNA or RNA (77% and 73%, respectively) were from outside the rectal mucosa. Cell-associated SIV RNA and DNA were not detected in any tissues from an SIV-negative macaque, but were readily detected in all tissues of macaques with conventionally controlled SIV infection. Overall, tissue levels in these conventional controllers averaged >3 logs higher than the measurable values of RhCMV/SIV-vector-vaccinated controllers ($P < 0.0001$ by the Wilcoxon rank sum test). Levels of cell-associated SIV were higher still in a macaque with poorly controlled infection. We also assayed lymphoid tissue cells from these macaques for the presence of inducible, replication-competent SIV by co-culture (Supplementary Table 1). All co-cultures (up to 20 replicates per specimen) from RhCMV/SIV-vector-vaccinated controllers were negative for recoverable SIV, whereas replication-competent SIV was readily detected in co-cultures of tissue cells from the conventional controllers. The paucity of SIV nucleic acid and the lack of recoverable SIV in RhCMV/SIV-vector-vaccinated controller macaques are in sharp contrast to the levels of HIV or SIV found in either humans or macaques receiving highly active antiretroviral therapy or in elite controllers^{16–21}, and suggest an unprecedented level of SIV control and even the possibility of progressive clearance of the SIV infection over time. Importantly, despite little or no SIV replication in the RhCMV/SIV-vector-vaccinated controllers, peripheral blood T cells specific for SIV proteins included in the RhCMV/SIV vectors (for example, Gag and Pol) were stably maintained at high frequency for 700 days after infection (CD8⁺ T-cell responses with $94 \pm 0.5\%$ T_{EM} phenotype); in

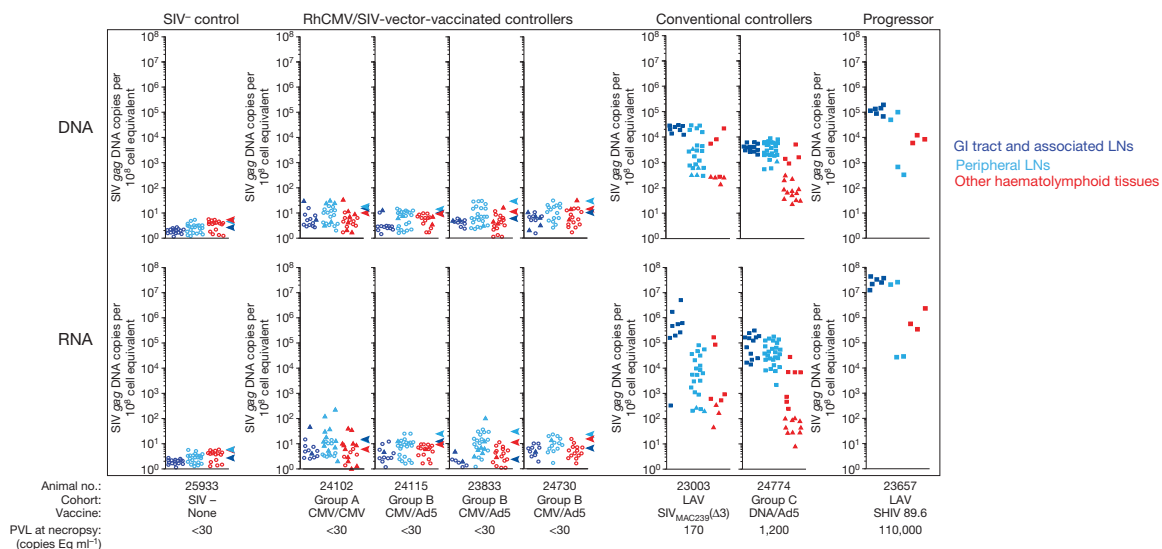


Figure 4 | Measurement of SIV RNA and DNA in long-term RhCMV/SIV-vector-vaccinated controllers. Nested quantitative PCR and quantitative RT-PCR analysis of SIV DNA and RNA, respectively, on tissue obtained at necropsy from an uninfected rhesus macaque, four long-term (>52 weeks) RhCMV/SIV-vector-vaccinated controller rhesus macaques (one Group A; three Group B), two conventional controller rhesus macaques (a live attenuated SIV (LAV)-vaccinated macaque that resisted wild-type SIV_{MAC239} challenge 33 and 10 weeks before necropsy, a Group C, DNA/Ad5-vaccine-protected macaque at 55 weeks post-infection), and a rhesus macaque with poorly controlled SIV infection (a LAV-vaccinated macaque 24 weeks after wild-type SIV_{MAC239} challenge). Filled square plot symbols indicate DNA or RNA copy numbers based on directly measured values for samples giving 10/10 replicate reactions positive. Filled triangles indicate results for samples giving at least one, but less

than ten, replicate reactions positive, with copy number imputed by Poisson distribution. Open circles indicate specimens that gave 0/10 replicates positive with the symbol's position in the plots at the threshold value corresponding to a Poisson distribution imputed copy number corresponding to 1/10 replicates positive. PVL, plasma viral load. All values are normalized for nucleic acid input. Arrowheads indicate the highest threshold value for negative samples (0/10 replicates positive) for all of the tissues analysed for that macaque. Gastrointestinal (GI) tract and associated lymph nodes (LNs) include colon/rectum, ileum, jejunum, superior/medial/inferior mesenteric and ileocaecal lymph nodes. Peripheral LNs include axillary, submandibular, inguinal, iliosacral and tracheobronchial lymph nodes. Other haematolymphoid tissues include liver, spleen, bone marrow, tonsil and thymus.

contrast to the SIV-infection-elicited Vif-specific responses; Fig. 3b and Supplementary Fig. 15). Thus, persistent RhCMV/SIV vectors provide for long-term maintenance of high-frequency SIV-specific T_{EM} responses, which would otherwise wane with stringent virological control, thereby ensuring continuous, high-level surveillance for SIV-infected cells, even when only rare infected cells are present.

In summary, the 16 long-term RhCMV/SIV-vector-vaccinated controllers described in this and our previous study⁵ unequivocally demonstrate a previously undescribed form of immune-mediated control of highly pathogenic SIV in which mucosally acquired infection is arrested before irreversible establishment of disseminated, progressive infection. Although stringently controlled, residual SIV infection is still present for weeks to months in most of these controllers, but wanes over time until eventually it is barely detectable by the most sensitive molecular virological and immunological criteria. The available data strongly indicate that this unique control is related to the high-frequency $CD8^{+}$, and possibly $CD4^{+}$, T_{EM} -biased, SIV-specific T-cell responses that are elicited and indefinitely maintained by the persistent RhCMV/SIV vectors, are situated in both mucosal portals of entry and potential sites of distant viral spread, and can protect without anamnestic expansion (see Supplementary Discussion). The ability of RhCMV/SIV vectors to indefinitely maintain SIV-specific T_{EM} responses in these sites, independent of the level of SIV replication, provides for continuous surveillance for SIV-infected cells, preventing relapse and, perhaps, ultimately clearing residual infection. Thus, CMV vectors provide a powerful new approach for HIV/AIDS vaccine development that could be used alone or in combination with complementary vaccine strategies that exploit different HIV immune vulnerabilities.

METHODS SUMMARY

Sixty-seven purpose-bred male rhesus macaques (*Macaca mulatta*) of Indian genetic descent were used in this study. RhCMV/SIV vectors were given subcutaneously at a dose of 5×10^6 plaque-forming units per vector. DNA and Ad5 vectors were given intramuscularly at doses of 1.6 mg per vector and 2×10^{10} particle units per vector, respectively. Rhesus macaques were challenged intrarectally with SIV_{MAC239} using a repeated (weekly) limiting dose protocol⁵. After the onset of infection (plasma viral load ≥ 30 SIV RNA copy equivalents (Eq) per ml), macaques were followed weekly until onset of AIDS or a minimum of 224 days for progressive infection and 365 days for controlled infection. SIV- and RhCMV-specific $CD4^{+}$ and $CD8^{+}$ T-cell responses were measured in mononuclear cell preparations from blood and tissues by flow cytometric intracellular cytokine analysis⁵. SIV Env-specific antibodies were determined by neutralization of tissue-culture-adapted SIV_{MAC251} using a luciferase reporter gene assay²². Levels of SIV RNA and DNA in plasma and from isolated cell preparations were quantified by standard quantitative real-time PCR and RT-PCR assays^{23,24}. Tissue-associated SIV RNA and DNA at necropsy were quantified by an ultra-sensitive nested, quantitative real-time PCR and RT-PCR approach (see Methods). The presence of inducible, replication-competent SIV in mononuclear cell preparations was detected by co-cultivation with CEMx174 cells, as previously described¹⁶.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 December 2010; accepted 17 March 2011.

Published online 11 May 2011.

- Grovit-Ferbas, K., Pappas, T. & O'Brien, W. A. in *Persistent Viral Infections* (eds Ahmed, R. & Chen, A. I.) 3–45 (John Wiley & Sons, 1999).
- Haase, A. T. Perils at mucosal front lines for HIV and SIV and their hosts. *Nature Rev. Immunol.* **5**, 783–792 (2005).
- Goulder, P. J. & Watkins, D. I. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nature Rev. Immunol.* **8**, 619–630 (2008).
- Haase, A. T. Targeting early infection to prevent HIV-1 mucosal transmission. *Nature* **464**, 217–223 (2010).
- Hansen, S. G. *et al.* Effector memory T cell responses are associated with protection of rhesus monkeys from mucosal simian immunodeficiency virus challenge. *Nature Med.* **15**, 293–299 (2009).
- Robinson, H. L. & Amara, R. R. T cell vaccines for microbial infections. *Nature Med.* **11**, S25–S32 (2005).
- Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl Acad. Sci. USA* **105**, 7552–7557 (2008).

- Chun, T. W. *et al.* Early establishment of a pool of latently infected, resting $CD4^{+}$ T cells during primary HIV-1 infection. *Proc. Natl Acad. Sci. USA* **95**, 8869–8873 (1998).
- Bachmann, M. F., Kundig, T. M., Hengartner, H. & Zinkernagel, R. M. Protection against immunopathological consequences of a viral infection by activated but not resting cytotoxic T cells: T cell memory without “memory T cells”? *Proc. Natl Acad. Sci. USA* **94**, 640–645 (1997).
- Ochsenbein, A. F. *et al.* A comparison of T cell memory against the same antigen induced by virus versus intracellular bacteria. *Proc. Natl Acad. Sci. USA* **96**, 9293–9298 (1999).
- Casimiro, D. R. *et al.* Attenuation of simian immunodeficiency virus SIVmac239 infection by prophylactic immunization with dna and recombinant adenoviral vaccine vectors expressing Gag. *J. Virol.* **79**, 15547–15555 (2005).
- Letvin, N. L. *et al.* Preserved $CD4^{+}$ central memory T cells and survival in vaccinated SIV-challenged monkeys. *Science* **312**, 1530–1533 (2006).
- Wilson, N. A. *et al.* Vaccine-induced cellular immune responses reduce plasma viral concentrations after repeated low-dose challenge with pathogenic simian immunodeficiency virus SIVmac239. *J. Virol.* **80**, 5875–5885 (2006).
- Friedrich, T. C. *et al.* Subdominant $CD8^{+}$ T-cell responses are involved in durable control of AIDS virus replication. *J. Virol.* **81**, 3465–3476 (2007).
- Reynolds, M. R. *et al.* Macaques vaccinated with live-attenuated SIV control replication of heterologous virus. *J. Exp. Med.* **205**, 2537–2550 (2008).
- Shen, A. *et al.* Resting $CD4^{+}$ T lymphocytes but not thymocytes provide a latent viral reservoir in a simian immunodeficiency virus-Macaca nemestrina model of human immunodeficiency virus type 1-infected patients on highly active antiretroviral therapy. *J. Virol.* **77**, 4938–4949 (2003).
- Dinso, J. B. *et al.* A simian immunodeficiency virus-infected macaque model to study viral reservoirs that persist during highly active antiretroviral therapy. *J. Virol.* **83**, 9247–9257 (2009).
- North, T. W. *et al.* Viral sanctuaries during highly active antiretroviral therapy in a nonhuman primate model for AIDS. *J. Virol.* **84**, 2913–2922 (2010).
- Ruiz, L. *et al.* Protease inhibitor-containing regimens compared with nucleoside analogues alone in the suppression of persistent HIV-1 replication in lymphoid tissue. *AIDS* **13**, F1–F8 (1999).
- Blankson, J. N. *et al.* Isolation and characterization of replication-competent human immunodeficiency virus type 1 from a subset of elite suppressors. *J. Virol.* **81**, 2508–2518 (2007).
- Anton, P. A. *et al.* Multiple measures of HIV burden in blood and tissue are correlated with each other but not with clinical parameters in aviremic subjects. *AIDS* **17**, 53–63 (2003).
- Montefiori, D. C. Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. *Curr. Protoc. Immunol.* **12**, Unit 12.11 (2005).
- Cline, A. N., Bess, J. W., Piatak, M. Jr & Lifson, J. D. Highly sensitive SIV plasma viral load assay: practical considerations, realistic performance expectations, and application to reverse engineering of vaccines for AIDS. *J. Med. Primatol.* **34**, 303–312 (2005).
- Venneti, S. *et al.* Longitudinal *in vivo* positron emission tomography imaging of infected and activated brain macrophages in a macaque model of human immunodeficiency virus encephalitis correlates with central and peripheral markers of encephalitis and areas of synaptic degeneration. *Am. J. Pathol.* **172**, 1603–1616 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the National Institute of Allergy and Infectious Diseases (RO1 AI060392; contract #HHSN272200900037C); the International AIDS Vaccine Initiative (IAVI) and its donors, particularly the United States Agency for International Development (USAID); the Bill & Melinda Gates Foundation-supported Collaboration for AIDS Vaccine Discovery; the National Center for Research Resources (P51 RR00163; R24 RR016001); and the National Cancer Institute (contract HHSN261200800001E). We thank A. Sylwester, D. Seiss, R. Lum, H. Park and A. Okoye for specialized technical assistance; P. Barry, G. Pavlakakis, G. Franchini, C. Miller, N. Wilson, and K. Reimann and Nonhuman Primate Reagent Resource for provision of crucial constructs or reagents; D. Watkins for MHC typing; D. Montefiori for neutralizing antibody assays; N. Letvin and L. Shen for TRIM5a typing; S. Mongoué-Tchokote and M. Mori for statistical assistance; A. Townsend and T. Schroyer for figure preparation; and K. Früh, D. Watkins, B. Beresford, A. McDermott, R. King and W. Koff for discussion and advice.

Author Contributions S.G.H. planned and performed experiments and analysed data, assisted by J.C.F., M.S.L., A.B.V., C.M.H., L.C.-J. and N.W. T.S., A.W.L. and M.K.A. managed the animal protocols. M.A.J. designed, constructed and characterized the RhCMV vectors. M.P. Jr and J.D.L. planned and performed SIV quantification studies, assisted by K.O. and R.S. C.L.P. and M.J.C. designed and constructed the DNA and Ad5 vectors used in this study. J.A.N. was involved in conception of the RhCMV vector strategy. L.J.P. conceived the RhCMV vector strategy, supervised experiments, analysed data and wrote the paper, assisted by S.G.H., M.A.J. and J.D.L.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interest: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.J.P. (pickerl@ohsu.edu).

METHODS

Rhesus macaques. A total of 68 purpose-bred juvenile and adult male rhesus macaques (*Macaca mulatta*) of Indian genetic descent were used in the experiments reported in this study, including 61 RhCMV⁺ macaques in the vaccination/challenge experiment shown in Fig. 1a (Group A and B: $n = 12$ each; Group C: $n = 9$; and Group D: $n = 28$; 11 concurrent and 17 historical controls), 4 long-term RhCMV/Gag-vector-vaccinated macaques, 2 live attenuated SIV-vector-vaccinated/wild-type SIV_{MAC239}-challenged macaques (one controller and one non-controller), and one unvaccinated, uninfected macaque. All macaques were free of cercopithecine herpesvirus 1, D-type simian retrovirus, simian T-lymphotrophic virus type 1 and SIV infection at the start of the study. Group A–C and concurrent Group D included 4 macaques each with the Mamu A*01 allele, but no macaques with the B*08 and B*17 alleles associated with post-peak control of SIV replication⁴. Of the 18 historical Group D rhesus macaques, 11 lacked A*01, B*08 and B*17 alleles, and 6 expressed either B*08 or B*17. The latter macaques were excluded from analysis of viral load in progressive infection, as they expressed protective alleles that were not represented in the vaccine groups. Rhesus macaques were used with approval of the Oregon National Primate Research Center Animal Care and Use Committee under the standards of the US National Institutes of Health Guide for the Care and Use of Laboratory Animals. RhCMV/SIV vectors (RhCMV Gag, Rev/Nef/Tat, Env, Pol1 and Pol2, see later) were given subcutaneously at a dose of 5×10^6 plaque-forming units. Ad5-SIV vectors (Ad5 Gag, Env, Pol, Nef and VVVTR, see later) were given intramuscularly (i.m.) at a dose of 2×10^{10} particle units per vector, and DNA (i.m. at 1.6 mg per vector) was given at weeks 0, 4 and 8 for the DNA prime animals before Ad5-SIV boost at week 14. Plasmids expressing a fusion protein comprised of SIV Vif, Vpr, Vpx, Tat and Rev, or individual Gag, Pol, Env and Nef open reading frames (ORFs) were used for DNA priming. Rhesus macaques were challenged intrarectally with highly pathogenic SIV_{MAC239} using the repeated (weekly), limiting dose protocol described previously⁵. Plasma viral loads were measured weekly, with challenge discontinued the week after detection of >30 copy equivalents (Eq) per ml of SIV RNA (with the challenge preceding the first measured plasma viral load of >30 copy Eq ml⁻¹ considered the day of infection). Macaques were considered controllers if plasma viral load became undetectable within 2 weeks of the initial positive plasma viral load and was then maintained below threshold for at least 4 of the subsequent 5 weeks. Challenged macaques were followed until onset of AIDS or a minimum of 224 days for progressive infection and a minimum of 365 days for controlled infection. For CD8⁺ lymphocyte depletion, rhesus macaques were administered 10, 5, 5 and 5 mg per kg body weight of the humanized monoclonal anti-CD8 α antibody cM-T807, on days 0, 3, 7 and 10, respectively²⁵. For CD4⁺ lymphocyte depletion, rhesus macaques were administered one dose of the humanized monoclonal anti-CD4 antibody huOKT4A at 50 mg per kg body weight²⁶. Mononuclear cell preparations were obtained from blood, bone marrow, bronchoalveolar lavage (BAL), lymph nodes, spleen, liver, tonsil, thymus and intestinal mucosa as previously described^{27–29}. For SIV quantification by nested qPCR/RT-PCR, whole tissue pieces obtained at necropsy were then flash frozen in liquid nitrogen and stored at -80°C before nucleic acid isolation.

RhCMV/SIV vectors. Construction and characterization of RhCMV Gag, RhCMV Retanef and RhCMV Env has been described⁵. Two additional RhCMV viruses expressing either the 5' (protease/reverse transcriptase; designated RhCMV Pol1), or 3' (RNase H/integrase; designated RhCMV Pol2) of SIV_{MAC239} Pol were constructed in an identical fashion by using E/T recombination and the RhCMV (68–1) bacterial artificial chromosome (BAC) (pRhCMV/BAC-Cre)³⁰. Deletions were introduced into Pol to inactivate protease ($\Delta 25$ -DTG-27), reverse transcriptase ($\Delta 184$ -YMDD-187), RNaseH ($\Delta E478$) and integrase ($\Delta D64$, $\Delta D116$, and $\Delta E152$)³¹. Pol protein expression was placed under control of the EF1 α promoter to achieve maximal expression. In all RhCMV/SIV vectors, the SIV antigen-expressing cassettes are inserted into the pRhCMV/BAC-Cre at nucleotide 207,630 within a non-coding region between rh213 and Rh214. RhCMV/SIV viruses were reconstituted by transfection of recombinant BAC DNA into RhCMV-permissive macaque fibroblasts. Following virus reconstitution and BAC cassette 'self-excision', RhCMV/SIV vectors contain the entire wild-type (68–1) RhCMV genome³⁰. Vector SIV antigen expression was confirmed by western blot analysis using antibodies to Flag (Sigma-Aldrich; RhCMV Gag), V5 (Invitrogen; RhCMV Retanef), c-Myc/KK45 (Sigma-Aldrich; RhCMV Env), and HA (Sigma-Aldrich; RhCMV Pol1 and Pol2).

DNA vaccines. The plasmid vaccine immunogens used in this study covered the full SIV_{MAC239} (NCBI M33261) genome. ORFs were sequence-optimized for expression in mammalian cells and cloned into a plasmid DNA expression vector. Expression of the intronless coding sequences was controlled by the human CMV (HCMV) promoter/enhancer and the bovine growth hormone polyadenylation signal³². Gag (12S), Pol (91S), Env (99S) and Nef (pCMV-Nef) were expressed as single polypeptides, whereas Vif, Vpr, Vpx, Tat and Rev were expressed as a fusion protein (pCMV-VVVTR). The 12S Gag plasmid expresses a myristoylated Gag

protein spanning amino acids (aa) 1–508 and lacks two C-terminal aa, but is otherwise similar to a Gag plasmid reported previously³³. Plasmid 99S expresses a native form of gp160, as previously reported^{31,33}. Sequence-optimized Nef without a myristoylation signal was PCR-amplified from 179S plasmid³¹, and subsequently transferred into the pCMV vector. The Pol (91S) coding sequence contained deletion mutations to inactivate protease, reverse transcriptase, RNaseH and integrase, as described³¹. Large-scale plasmid production for immunization was prepared by Aldevron, LLC. Expression of plasmids after transient transfection of HEK 293 cells was corroborated by western blot using polyclonal antibodies against SIV_{MAC239} Gag, Env, Nef and Rev proteins made in-house, and anti-SIV_{MAC251} serum from the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH.

Adenovirus vectors. Plasmid SIV_{MAC239} sequences, again covering the full SIV_{MAC239} genome, were cloned into a human adenovirus serotype 5 (Ad5), which lacks E1A and most of E1B and $\Delta E3$ using the Adeasy Adenoviral vector system (Stratagene) to make the Ad5-SIV Gag, Ad5-SIV Env, Ad5-SIV Pol, Ad5-SIV Nef and Ad5-SIV VVVTR vectors. SIV_{MAC239} genes were inserted into the E1 region of the Ad5 under the control of the HCMV immediate early promoter/enhancer and the SV40 polyadenylation signal. All vectors were rescued and propagated in HEK 293 cells and purified by double cesium chloride centrifugation³⁴. Dosing was based on the physical number of particles (PU) of Ad5 as measured by spectrophotometry³⁵. Expression of SIV proteins from Ad5 vectors after A549 infection was confirmed by western blot as described earlier.

SIV_{MAC239} challenge virus. The pathogenic SIV_{MAC239} challenge virus stock (provided by C. Miller) was generated by expanding the SIV_{MAC239} clone in rhesus macaque peripheral blood mononuclear cells (PBMCs), and was quantified using the sMAGI cell assay and by quantitative RT-PCR for SIV genomic RNA⁵.

Immunological assays. SIV- and RhCMV-specific CD4⁺ and CD8⁺ T-cell responses were measured in mononuclear cell preparations from blood and tissues by flow cytometric intracellular cytokine analysis, as previously described in detail⁵. Briefly, sequential 15-mer peptides (overlapping by 11 amino acids) comprising the SIV_{MAC239} Gag, Rev/Nef/Tat, Env, Pol, Vif and Vpr/Vpx proteins were used in the presence of co-stimulatory CD28 and CD49d monoclonal antibodies (BD Biosciences). Cells were incubated with antigen and co-stimulatory molecules alone for 1 h, followed by addition of Brefeldin A (Sigma-Aldrich) for an additional 8 h. Co-stimulation without antigen served as a background control. Cells were then stained with fluorochrome-conjugated monoclonal antibodies, flow cytometric data collected on an LSR II (BD Biosciences) and data analysed using the FlowJo software program (version 8.8.6; Tree Star). Response frequencies (CD69⁺/TNF⁺ and/or CD69⁺/IFN- γ ⁺) were first determined in the overall CD4⁺ and CD8⁺ population and then memory corrected (with memory T-cell subset populations delineated on the basis of CD28 and CD95 expression), as previously described^{5,27}. The data presented as 'end of vaccine phase' response frequencies represent an average of values obtained from samples collected on days 379, 392, 401 and 413 after initial vaccination. Titres of SIV Env-specific antibodies were determined by neutralization of tissue culture-adapted SIV_{MAC251} using a luciferase reporter gene assay²².

Viral detection assays. Quantitative real-time RT-PCR and PCR assays targeting a highly conserved sequence in Gag were used for standard measurements of plasma SIV RNA and cell-associated SIV RNA and DNA within peripheral blood and lymph node mononuclear cells, as previously described^{23,24}. For plasma testing of a sample to score as positive, duplicate amplification reactions yielding ≥ 30 copy Eq ml⁻¹ were required. Isolated viral blips were also repeated in a duplicate sample in almost all instances, and in macaques where infection was manifest only by isolated viral blips (Group A and B controllers), infection was confirmed by the development (Group A) or boosting (Group B) of T-cell responses specific for SIV Vif, an SIV antigen not included in the RhCMV/SIV vectors (Supplementary Fig. 12), as described⁵. To further address the possibility of false positive amplification reactions, we analysed 136 known SIV-negative (pre-challenge) samples from the present study, and from other studies from the same facility and investigators and others. Samples were run in the same laboratory, over approximately the same period of time, using the same procedures, with negative specimens interspersed with positive samples in assay runs. Zero of 136 known SIV-negative samples scored positive by the above criteria, which was significantly different from 84 positive samples (viral 'blips') of 678 total samples (12.4%) obtained from RhCMV/SIV-vector-vaccinated controllers in vaccine Groups A and B ($P < 0.0001$, Fisher's exact test). To more precisely characterize levels of SIV DNA and RNA from tissues of RhCMV/SIV-vector-vaccinated controllers (which were below the sensitivity threshold of the standard assays), we used a new ultra-sensitive, nested, quantitative real-time PCR/RT-PCR method. Tissue pieces were collected directly into extraction tubes and immediately frozen using liquid nitrogen. Samples were stored at -80°C and handled on dry ice until stabilized in extraction solution. Specimens of approximately 100 mg or less were homogenized in 1 ml of TriReagent (Molecular Research

Center) in 2 ml extraction tubes of Lysing Matrix D using FastPrep instrumentation (MP Biomedicals) according to the manufacturer's recommendations. Total RNA and DNA were prepared from the homogenates following manufacturer's recommendations, but specifically following the alternative, back-extraction method for DNA extraction. Recovered RNA and DNA were dissolved in minimal volumes of 10 mM TrisCl, pH 8.0 and 10 mM TrisCl, pH 9.0, respectively, as appropriate for replicate testing in qRT-PCR and qPCR protocols. Tissue specimens greater than 100 mg in mass were initially pulverized to powder under cryogenic conditions before extraction of RNA and DNA. Pulverization was accomplished in 15 ml polycarbonate extraction tubes with stainless steel grinding balls using rapid vertical shaking, all being maintained at appropriate temperatures in aluminium blocks pre-chilled in liquid N₂ (GenoGrinder, SPEX SamplePrep). Pulverized tissue powder was then suspended in 3–10 ml of TriReagent, depending on the starting amount of tissue. Total RNA and DNA were then prepared from 1 ml of TriReagent suspension as described earlier; residual suspension was archived at –80 °C for additional analysis, as necessary. To maximize sensitivity, nested quantitative real-time PCR/RT-PCR protocols were designed to accommodate higher amounts of input nucleic acid, and potential inhibitors, than are typically tolerated in standard assays. Reaction conditions and thermal profiles followed those referenced above for the plasma and isolated cell assays^{23,24} with two exceptions: (1) in the quantitative RT-PCR assay, the 'nested' reverse primer, as opposed to random hexamers, was used to prime cDNA synthesis specifically for SIV sequence, thereby avoiding generation of non-specific targets and further enhancing the direct sensitivity of detection of SIV RNA; and (2) 2.5 units of PlatinumTaq (Invitrogen), rather than 1.25 units of TaqGold (Applied Biosystems), were used in the amplification steps. A 'nested' or 'pre-amplification' of cDNA or DNA was performed for 12 cycles with the application of primers, SIVnestF01 (GATTTGGATTAGCAG AAAGCCTGTTG) and SIVnestR01 (GTTG GTCTACTTGTTTGGCATAGTTTC), flanking the SIV Gag target region. Five microlitres of this first amplification were then transferred to 50 microlitres of cocktail for amplification of the SIV gag DNA target sequence in duplex with amplification of a single copy rhesus CCR5 target sequence for normalization, as referenced earlier^{23,24}. Real-time PCR was then performed. For both RNA and DNA determinations, 12 replicate reactions were tested per sample including a spike of 10 copies of RNA or DNA internal control sequence standard in two of the 12 reactions to assess overall amplification efficiency and assess potential inhibition of the PCR or RT-PCR. Samples showing greater than a 5 cycle shift in amplification of the spiked standard, compared to amplification in the absence of specimen nucleic acid, corresponding to less than 74% overall amplification efficiency, were diluted and re-assayed. Quantitative determinations for samples showing amplification in all replicates were derived directly with reference to a standard curve. Quantitative determinations for samples showing less than 10 positive amplifications in replicates were derived from the frequency of positive amplifications, corresponding to the presence of at least one target copy in a reaction, according to a Poisson distribution of a given median copy number per reaction. It should be noted that this assay yielded no positive reactions out of a total of 1,100 total reactions (RNA and DNA) from tissues derived from an SIV-uninfected rhesus macaque, which was significantly different from 178 positive reactions of 4,310 total reactions (4.1%) from tissues derived from the 4 RhCMV/SIV-vector-vaccinated controllers studied at

necropsy ($P < 0.0001$; Fisher's exact test). The presence of inducible, replication competent SIV in mononuclear cell preparations derived from different tissue sites at necropsy was detected by co-cultivation with CEMx174 cells, as previously described^{16,36}. To detect shedding of RhCMV/SIV vectors in the urine of vaccinated rhesus macaques, virus was concentrated from cleared pan-collected urine and co-cultured with macaque fibroblasts. Cell lysates were collected after development of cytopathic effect or after 28 days, and assessed for vector replication based on expression of SIV antigen-specific epitope tags by western blot analysis⁵.

Statistical analysis. We performed statistical analysis with SAS version 9.1 (Statistical Analysis System). Individual tests are described in the figure legends for all analyses. Briefly, Fisher's exact test was used to determine significance of categorical data such as the fraction of controllers versus non-controllers in the different vaccine groups. The Wilcoxon rank sum and Kruskal-Wallis tests were used to compare populations of continuous data for groups of 2 and ≥ 3 , respectively. Analysis of variance and a test for linear trend were used to determine the significance of the reduction in viral blip frequency over time in RhCMV-vector-vaccinated controllers. In all analyses, we used a two-sided significance level (α) of 0.05, with correction made for multiple comparisons using the Bonferroni method.

25. Okoye, A. *et al.* Profound CD4⁺/CCR5⁺ T cell expansion is induced by CD8⁺ lymphocyte depletion but does not account for accelerated SIV pathogenesis. *J. Exp. Med.* **206**, 1575–1588 (2009).
26. Vaccari, M. *et al.* Reduced protection from simian immunodeficiency virus SIVmac251 infection afforded by memory CD8⁺ T cells induced by vaccination during CD4⁺ T-cell deficiency. *J. Virol.* **82**, 9629–9638 (2008).
27. Pitcher, C. J. *et al.* Development and homeostasis of T cell memory in rhesus macaque. *J. Immunol.* **168**, 29–43 (2002).
28. Veazey, R. S. *et al.* Gastrointestinal tract as a major site of CD4⁺ T cell depletion and viral replication in SIV infection. *Science* **280**, 427–431 (1998).
29. Schmitz, J. E. *et al.* Simian immunodeficiency virus (SIV)-specific CTL are present in large numbers in livers of SIV-infected rhesus monkeys. *J. Immunol.* **164**, 6015–6019 (2000).
30. Chang, W. L. & Barry, P. A. Cloning of the full-length rhesus cytomegalovirus genome as an infectious and self-excisable bacterial artificial chromosome for analysis of viral pathogenesis. *J. Virol.* **77**, 5073–5083 (2003).
31. Kulkarni, V. R. *et al.* Comparison of immune responses generated by optimized DNA vaccination against SIV antigens in mice and macaques. *Vaccine* doi:10.1016/j.vaccine.2010.12.056.
32. Rosati, M. *et al.* DNA vaccines expressing different forms of simian immunodeficiency virus antigens decrease viremia upon SIVmac251 challenge. *J. Virol.* **79**, 8480–8492 (2005).
33. Rosati, M. *et al.* DNA vaccination in rhesus macaques induces potent immune responses and decreases acute and chronic viremia after SIVmac251 challenge. *Proc. Natl Acad. Sci. USA* **106**, 15831–15836 (2009).
34. Rosenfeld, M. A. *et al.* *In vivo* transfer of the human cystic fibrosis transmembrane conductance regulator gene to the airway epithelium. *Cell* **68**, 143–155 (1992).
35. Mittereder, N., March, K. L. & Trapnell, B. C. Evaluation of the concentration and bioactivity of adenovirus vectors for gene therapy. *J. Virol.* **70**, 7498–7509 (1996).
36. Shen, A. *et al.* Novel pathway for induction of latent virus from resting CD4⁺ T cells in the simian immunodeficiency virus/macaque model of human immunodeficiency virus type 1 latency. *J. Virol.* **81**, 1660–1670 (2007).

A graphene-based broadband optical modulator

Ming Liu^{1*}, Xiaobo Yin^{1*}, Erick Ulin-Avila¹, Baisong Geng², Thomas Zentgraf¹, Long Ju², Feng Wang^{2,3} & Xiang Zhang^{1,3}

Integrated optical modulators with high modulation speed, small footprint and large optical bandwidth are poised to be the enabling devices for on-chip optical interconnects^{1,2}. Semiconductor modulators have therefore been heavily researched over the past few years. However, the device footprint of silicon-based modulators is of the order of millimetres, owing to its weak electro-optical properties³. Germanium and compound semiconductors, on the other hand, face the major challenge of integration with existing silicon electronics and photonics platforms^{4–6}. Integrating silicon modulators with high-quality-factor optical resonators increases the modulation strength, but these devices suffer from intrinsic narrow bandwidth and require sophisticated optical design; they also have stringent fabrication requirements and limited temperature tolerances⁷. Finding a complementary metal-oxide-semiconductor (CMOS)-compatible material with adequate modulation speed and strength has therefore become a task of not only scientific interest, but also industrial importance. Here we experimentally demonstrate a broadband, high-speed, waveguide-integrated electroabsorption modulator based on monolayer graphene. By electrically tuning the Fermi level of the graphene sheet, we demonstrate modulation of the guided light at frequencies over 1 GHz, together with a broad operation spectrum that ranges from 1.35 to 1.6 μm under ambient conditions. The high modulation efficiency of graphene results in an active device area of merely 25 μm^2 , which is among the smallest to date. This graphene-based optical modulation mechanism, with combined advantages of compact footprint, low operation voltage and ultrafast modulation speed across a broad range of wavelengths, can enable novel architectures for on-chip optical communications.

Graphene, a single sheet of carbon atoms in a hexagonal lattice, has attracted great interest because of its exceptional electrical and optical properties^{8–13}. Graphene couples strongly to light, which enables observation of monolayer graphene under an optical microscope with the naked eye. A pristine graphene monolayer has a constant absorption of $\pi\alpha = 2.3\%$ across the infrared and visible range, where α is the fine-structure constant ($e^2/\hbar c$, where e is the electronic charge, \hbar is Planck's constant divided by 2π , and c the velocity of light). Moreover, the broad optical absorption in graphene can be controlled through electrical gating: by shifting the electronic Fermi level, one can controllably change graphene's optical transitions^{14,15}. The strong electroabsorption effect, which has not yet been observed in bulk materials, originates from graphene's unique electronic structure and its two-dimensional character. It implies that graphene has the potential to be used as the active medium in an optical electroabsorption modulator. However, one of the challenges involved in a direct graphene modulator is the limited absorption of a monolayer. This can be overcome by integrating graphene with an optical waveguide, which greatly increases the interaction length through the coupling between the evanescent waves and graphene.

A graphene-based waveguide-integrated electroabsorption modulator has several distinctive advantages. (1) Strong light–graphene interaction. In comparison to compound semiconductors, such as

those exhibiting the quantum-well with quantum-confined Stark effect (QCSE)⁶, a monolayer of graphene possesses a much stronger inter-band optical transition, which finds applications in novel optoelectronic devices such as photodetectors^{16,17}. (2) Broadband operation. As the high frequency dynamic conductivity for Dirac fermions is constant, the optical absorption of graphene is independent of wavelength, covering all telecommunications bandwidth and also the mid- and far-infrared^{18,19}. (3) High-speed operation. With a carrier mobility exceeding 200,000 $\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$ at room temperature^{20,21} (this is among the highest known), the Fermi level and hence the optical absorptions of graphene can be rapidly modulated through the band-filling effect. In addition, speed limiting processes in graphene (such as photocarrier generation and relaxation) operate on the timescale of picoseconds²², which implies that graphene-based electronics may have the potential to operate at 500 GHz, depending on the carrier density and graphene quality. (4) Compatibility with CMOS processing. The athermal optoelectronic properties of graphene and its CMOS-compatible integration processes at wafer scale^{23–25} make it a promising candidate for post-CMOS electronics, particularly for high frequency applications. With all these advantages, monolithic integration of a graphene electroabsorption modulator could open new routes to integrated photonics, with a compact footprint, low voltage operation and ultrafast modulation across a broad range of wavelengths.

Here we report the first waveguide-integrated graphene-based electroabsorption modulator, in which modulation is achieved by actively tuning the Fermi level of a monolayer graphene sheet. The gigahertz graphene modulator demonstrates a strong electroabsorption modulation of 0.1 $\text{dB} \mu\text{m}^{-1}$ and operates over a broad range of wavelength, from 1.35 μm to 1.6 μm , under ambient conditions.

The structure of the electroabsorption modulator is schematically illustrated in Fig. 1. A 50-nm-thick Si layer was used to connect the 250-nm-thick Si bus waveguide and one of the gold electrodes. Both the silicon layer and the waveguide were shallowly doped with boron to reduce the sheet resistance. A spacer of 7-nm-thick Al_2O_3 was then uniformly deposited on the surface of the waveguide by atom layer deposition. A graphene sheet grown by chemical vapour deposition^{23,24,26} (CVD) was then mechanically transferred onto the Si waveguide. To further reduce the access resistance of the device, the counter electrode was extended towards the bus waveguide by depositing a platinum (10 nm) film on top of the graphene layer. The minimum distance between platinum electrode and waveguide was controlled at 500 nm, so that the optical modes of the waveguide remained undisturbed by the platinum contact. The excess graphene was removed by oxygen plasma, leaving only the regions on top of the waveguide and between the waveguide and the platinum electrode.

The cross-sectional view of the device structure and the optical field distribution of the guided mode are shown in Fig. 1b. The thin silicon layer and the platinum electrode adjacent to the waveguide have negligible effect on the mode profile. To further improve the electroabsorption modulation efficiency, the silicon waveguide was designed to have the electric field maximized at its top and bottom surfaces, so that the interband transitions in graphene are also maximized (Fig. 1b). As

¹NSF Nano-scale Science and Engineering Center (NSEC), 3112 Etcheverry Hall, University of California at Berkeley, Berkeley, California 94720, USA. ²Department of Physics, University of California at Berkeley, Berkeley, California 94720, USA. ³Materials Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA.

*These authors contributed equally to this work.

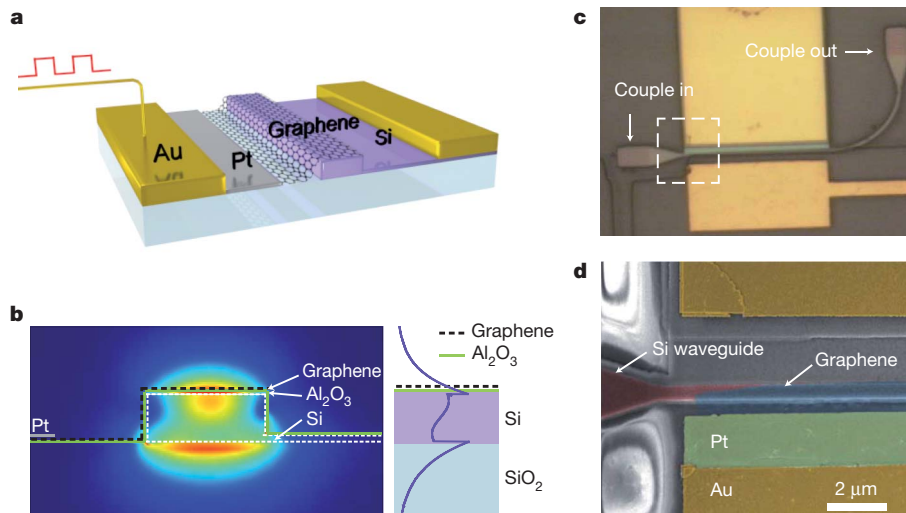


Figure 1 | A graphene-based waveguide-integrated optical modulator. **a**, Three-dimensional schematic illustration of the device; a monolayer graphene sheet is on top of a silicon bus waveguide, separated from it by a 7-nm-thick Al_2O_3 layer (not shown). The silicon waveguide is doped, and connected to the electrode (right, shown gold) through a thin layer of silicon defined by selective etching. **b**, Left, cross-section of the device, with an overlay of the optical mode plot, calculated by finite element simulation. The waveguide was carrying a single optical mode, and was designed so as to maximize the field at the interface between the waveguide and the graphene, to maximize the absorption efficiency. The thin silicon layer and the Pt electrode, which is 500 nm away from the waveguide, have negligible influence on the optical mode. Right, a cross-section through the centre of the waveguide; the purple

curve shows the magnitude of the electric field. The actual thicknesses of the graphene sheet and the Al_2O_3 in the simulation are 0.7 nm and 7 nm, respectively. **c**, Top-view optical microscope image of the waveguide. The Si waveguide was bent by 90° to change the polarization state between the input and the output light, to improve the signal-noise ratio. **d**, SEM image showing the boxed region in **c** at higher magnification, showing the detailed structure of the graphene modulator. False colours are used to highlight the Au electrode (yellow), the Pt electrode (Green), the graphene sheet (blue) and the waveguide (red). The width of the Si waveguide is 600 nm, while the distance between the Pt electrode and the Si waveguide is 500 nm. The bright multi-ring region beside the Au electrodes is due to a charging effect on the SiO_2 layer in the SEM.

graphene only interacts with the tangential (in-plane) electric field of electromagnetic waves, the graphene modulator is polarization-sensitive, as are conventional semiconductor-based electro-optical modulators³.

Figure 1c shows a top-view optical microscope image of the device, and a close-up scanning electron microscopy image of the active region is given in Fig. 1d. The graphene sheet, highlighted in Fig. 1d by a false blue colour, covers only the waveguide region; this is to minimize the capacitance. The platinum electrode (green) is placed 500 nm away from the 600-nm-wide Si waveguide. Light was coupled in and out of the waveguide through tapered gratings, which contribute most to the overall loss of the system. The Si waveguide was bent 90° to change the polarization state between the input and the output light, to improve the signal-noise ratio.

Figure 2 displays the transmission of $1.53\text{ }\mu\text{m}$ photons through the waveguide at different drive voltages, V_D . At low drive voltage ($-1\text{ V} < V_D < 3.8\text{ V}$), the Fermi level $E_F(V_D)$ of graphene is close to the Dirac point ($E_F(V_D) < \hbar\nu_0/2$), and interband transitions occur when electrons are excited by the incoming photons ($\hbar\nu_0$). The optical absorption of graphene is determined by the position of the Fermi level. By applying a drive voltage between the graphene and the waveguide, we can tune the Fermi level of the graphene, and therefore modulate the total transmission. With the current waveguide design, the modulation depth is as high as $0.1\text{ dB }\mu\text{m}^{-1}$, resulting in a graphene electroabsorption modulator with a footprint of merely $25\text{ }\mu\text{m}^2$. At large negative V_D ($< -1\text{ V}$), the Fermi level is lowered below the transition threshold ($E_F(V_D) = \hbar\nu_0/2$) owing to positive charge accumulation. As a result, there are no electrons available for interband transitions, and hence the graphene appears transparent. On the other hand, at large positive V_D ($> 3.8\text{ V}$), all electron states are filled up, and no interband transitions are allowed. Ideally, there should be a sharp change in transmission at $E_F(V_D) = \hbar\nu_0/2$. In reality, this transition was broadened owing to defects in the graphene, and shifted to higher voltage owing to natural doping from the substrate²⁷. When the modulator is in operation (that is, when no interband absorption is allowed),

its insertion loss is negligible as the intraband absorption of graphene is extremely low at near-infrared wavelengths²⁸.

To measure the dynamic response of the graphene modulator, radio frequency signals generated by a network analyser were added on a static V_D and applied to the modulator. The same $1.53\text{-}\mu\text{m}$ laser was used to test the modulator, and the out-coupled light was sent to a high-speed photodetector. Shown in Fig. 3 are the V_D -dependent r.f. responses of the graphene modulator; gigahertz operation of the device

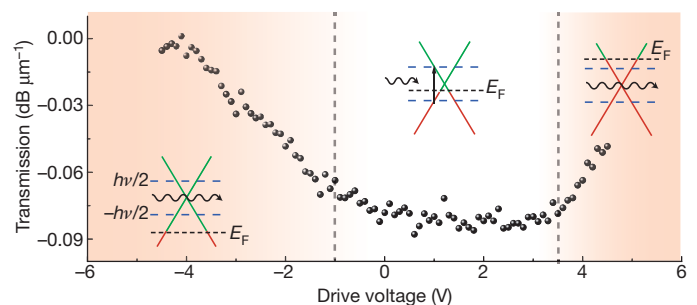


Figure 2 | Static electro-optical response of the device at different drive voltages. The main panel shows the modulation depth, normalized to the device length ($40\text{ }\mu\text{m}$), at different drive voltages (V_D). Three regions can be seen, and their band structures are shown as insets. In the middle region (with V_D in the range -1 V to 3.8 V), the Fermi level (E_F , black dashed line) is close to the Dirac point and the interband transition is allowed from electron-occupied regions (red lines) to unoccupied regions (green lines). Thus the graphene sheet is absorptive to incident photons ($\hbar\nu$), resulting in a modulation depth of $0.1\text{ dB }\mu\text{m}^{-1}$ and a miniaturized footprint of the modulator. In the left-hand region (with $V_D < -1\text{ V}$), the Fermi level (E_F) is lower than half the photon energy ($-\hbar\nu/2$, blue dashed line) and there are no electrons available for the interband transition. In the right-hand region (with $V_D > 3.8\text{ V}$), all electron states in resonance with incident photons ($\hbar\nu$) are occupied, and the transition is forbidden. In both of the last two cases, the transmission increases. The natural doping from the substrate offsets the centre of the absorption curve from zero bias. The transmission is measured at a laser wavelength of $1.53\text{ }\mu\text{m}$.

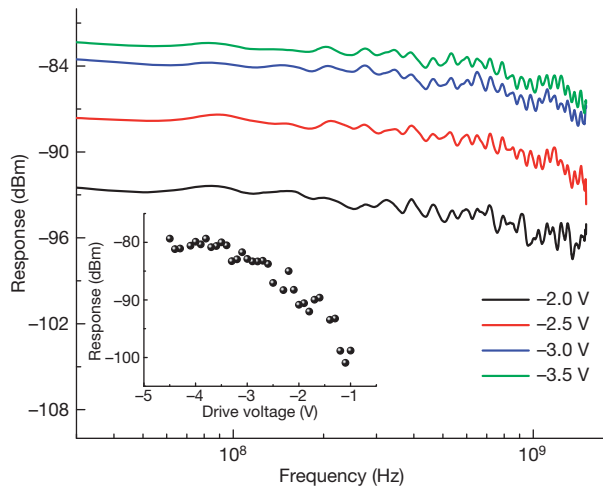


Figure 3 | Dynamic electro-optical response of the device. The main panel shows the response of the device as a function of frequency. Each curve corresponds to a different drive voltage: -2.0 V, -2.5 V, -3.0 V and -3.5 V (black, red, blue and green curve, respectively). The measured 3 dB bandwidths of the device are respectively 0.8 GHz, 1.1 GHz, 1.1 GHz and 1.2 GHz; the bandwidths are mainly restricted by the parasitic response of the device. Inset, low-frequency device response with different drive voltages, indicating that the device has best performance at a drive voltage of -4 V. The laser wavelength is 1.53 μm in the test.

at various drive voltages is obtained. Owing to the exceptionally high carrier mobility and saturation velocity of graphene, the bandwidth is not limited by the carrier transit time, but by the parasitic response of the device. With the platinum electrode placed 500 nm away from the waveguide, the total resistance of the system is reduced to around $600\ \Omega$. This resistance, together with the capacitance (of the order of 0.22 pF), limits the operation bandwidth of the present device to about 1 GHz (see Fig. 3 legend).

The device response at low frequency (300 kHz) is shown in Fig. 3 inset. At low V_D , the modulation response is weak, as the optical transmission is insensitive to V_D . When the drive voltage is increased, the r.f. response increases to a maximum at $V_D = -4$ V. As the drive voltage increases further, the modulation efficiency saturates, as graphene is then transparent within the modulation range of the bias voltage.

As the overall optical opacity of graphene is independent of wavelength and the high frequency dynamic conductivity for Dirac fermions is constant, the graphene electroabsorption modulator is therefore intrinsically broadband, unlike modulators that are based on optical cavities or resonant optical effects (such as QCSE). In order to investigate this broadband effect, we study the static response of the device to a white light source from a supercontinuous laser. The response is shown as a function of wavelength and V_D in Fig. 4a; we refer to this as a two-dimensional (2D) transmission spectrum. A 3 dB modulation, corresponding to a transmission value of 2 (a.u.) in Fig. 4a, is achieved for a broad band of wavelengths, from $1.35\ \mu\text{m}$ to $1.6\ \mu\text{m}$. Although a higher modulation depth and broader wavelength range are expected at a higher drive voltage, we chose to use low drive voltage not only to avoid spacer oxide breakdown but also because high drive voltages increase power consumption and violate voltage restrictions in CMOS devices.

Two-dimensional transmission spectra also allow us to determine the electronic band dispersion of the graphene. As the graphene electroabsorption modulation is dictated by the optical transition, $h\nu = 2E_F$, the graphene modulator has a different response at different wavelengths. Operation at higher photon energy ($h\nu$) always requires a larger change in the Fermi level (E_F). The trace of critical drive voltage (V) for maximum transmission change rate, shown as a dashed line in Fig. 4a, is defined by $h\nu = 2E_F = 2h\nu_F \sqrt{\eta\pi|V + V_0|}$, where ν_F is the Fermi velocity, V_0 is the voltage offset caused by natural doping, and

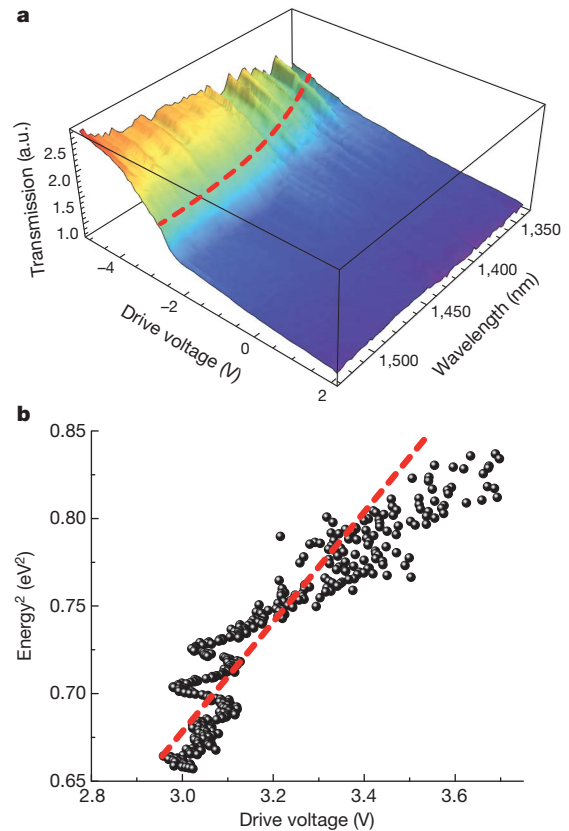


Figure 4 | Spectrum characterization of the optical modulator. **a**, The transmission of the device as a function of drive voltage and wavelength ($1,350$ – $1,600$ nm). The transmission is normalized to $V_D = 1$ V. The red dashed curve shows the trace for maximum transmission change rate, which unambiguously bends to higher drive voltages at shorter wavelengths. **b**, Squared photon energy versus critical drive voltage for maximum transmission change rate. The red dashed line shows a linear fit to the experimental data, which directly yields the Fermi velocity of $\nu_F = 0.9 \times 10^6$ m s $^{-1}$.

$\eta = 9 \times 10^{16}$ m $^{-2}$ V $^{-1}$, as estimated using a parallel-plate capacitor model of our device. The relation between critical drive voltage and the square of the photon energy is plotted in Fig. 4b. The linear fit (red dashed line) yields the voltage offset (-0.8 V) and the Fermi velocity (0.9×10^6 m s $^{-1}$), which agree with other reported values²⁹.

We have demonstrated a graphene-based optical modulator that has broad optical bandwidth (1.35 – $1.6\ \mu\text{m}$), small device footprint ($25\ \mu\text{m}^2$) and high operation speed (1.2 GHz at 3 dB) under ambient conditions, all of which are essential for optical interconnects for future integrated optoelectronic systems. The modulation efficiency of a single layer of carbon atoms in a hexagonal lattice (graphene) is already comparable to, if not better than, traditional semiconductor materials such as Si, GeSi and InGaAs, which are orders of magnitude larger in active volume. The flexibility of graphene sheets could also enable radically different photonic devices. For example, graphene could be integrated with flexible substrate and plastic waveguides³⁰. Or it could be used in novel geometries, such as a flexible modulator on a nano-optical cable. The recent development of large scale graphene synthesis and transfer techniques²³ ensure its compatibility with the existing integrated electronics platform.

Received 3 November 2010; accepted 24 March 2011.

Published online 8 May 2011.

1. Miller, D. A. B. Are optical transistors the logical next step? *Nature Photon.* **4**, 3–5 (2010).
2. Reed, G. T., Mashanovich, G., Gardes, F. Y. & Thomson, D. J. Silicon optical modulators. *Nature Photon.* **4**, 518–526 (2010).
3. Liu, A. S. et al. A high-speed silicon optical modulator based on a metal-oxide-semiconductor capacitor. *Nature* **427**, 615–618 (2004).

4. Kuo, Y. H. *et al.* Strong quantum-confined Stark effect in germanium quantum-well structures on silicon. *Nature* **437**, 1334–1336 (2005).
5. Liu, J. *et al.* Waveguide-integrated, ultralow-energy GeSi electro-absorption modulators. *Nature Photon.* **2**, 433–437 (2008).
6. Miller, D. A. B. *et al.* Band-edge electroabsorption in quantum well structures—the quantum-confined Stark-effect. *Phys. Rev. Lett.* **53**, 2173–2176 (1984).
7. Xu, Q., Schmidt, B., Pradhan, S. & Lipson, M. Micrometre-scale silicon electro-optic modulator. *Nature* **435**, 325–327 (2005).
8. Novoselov, K. S. Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
9. Geim, A. K. & Novoselov, K. S. The rise of graphene. *Nature Mater.* **6**, 183–191 (2007).
10. Schwierz, F. Graphene transistors. *Nature Nanotechnol.* **5**, 487–496 (2010).
11. Bonaccorso, F., Sun, Z., Hasan, T. & Ferrari, A. Graphene photonics and optoelectronics. *Nature Photon.* **4**, 611–622 (2010).
12. Liao, L. *et al.* High-speed graphene transistors with a self-aligned nanowire gate. *Nature* **467**, 305–308 (2010).
13. Avouris, P., Chen, Z. H. & Perebeinos, V. Carbon-based electronics. *Nature Nanotechnol.* **2**, 605–615 (2007).
14. Wang, F. *et al.* Gate-variable optical transitions in graphene. *Science* **320**, 206–209 (2008).
15. Li, Z. Q. *et al.* Dirac charge dynamics in graphene by infrared spectroscopy. *Nature Phys.* **4**, 532–535 (2008).
16. Xia, F. N., Mueller, T., Lin, Y. M., Valdes-Garcia, A. & Avouris, P. Ultrafast graphene photodetector. *Nature Nanotechnol.* **4**, 839–843 (2009).
17. Xu, X., Gabor, N. M., Alden, J. S., van der Zande, A. M. & McEuen, P. L. Photo-thermoelectric effect at a graphene interface junction. *Nano Lett.* **10**, 562–566 (2010).
18. Nair, R. R. *et al.* Fine structure constant defines visual transparency of graphene. *Science* **320**, 1308 (2008).
19. Mak, K. F. *et al.* Measurement of the optical conductivity of graphene. *Phys. Rev. Lett.* **101**, 196405 (2008).
20. Bolotin, K. I. *et al.* Ultrahigh electron mobility in suspended graphene. *Solid State Commun.* **146**, 351–355 (2008).
21. Du, X., Skachko, I., Barker, A. & Andrei, E. Y. Approaching ballistic transport in suspended graphene. *Nature Nanotechnol.* **3**, 491–495 (2008).
22. Kampfrath, T., Perfetti, L., Schapper, F., Frischkorn, C. & Wolf, M. Strongly coupled optical phonons in the ultrafast dynamics of the electronic energy and current relaxation in graphite. *Phys. Rev. Lett.* **95**, 187403 (2005).
23. Kim, K. S. *et al.* Large-scale pattern growth of graphene films for stretchable transparent electrodes. *Nature* **457**, 706–710 (2009).
24. Bae, S. *et al.* Roll-to-roll production of 30-inch graphene films for transparent electrodes. *Nature Nanotechnol.* **5**, 574–578 (2010).
25. Reina, A. *et al.* Large area, few-layer graphene films on arbitrary substrates by chemical vapor deposition. *Nano Lett.* **9**, 30–35 (2009).
26. Li, X. *et al.* Large-area synthesis of high-quality and uniform graphene films on copper foils. *Science* **324**, 1312–1314 (2009).
27. Zhang, Y. B., Brar, V. W., Girit, C., Zettl, A. & Crommie, M. F. Origin of spatial charge inhomogeneity in graphene. *Nature Phys.* **5**, 722–726 (2009).
28. Jablan, M., Buljan, H. & Soljačić, M. Plasmonics in graphene at infrared frequencies. *Phys. Rev. B* **80**, 245435 (2009).
29. Zhou, S. Y. *et al.* First direct observation of Dirac fermions in graphite. *Nature Phys.* **2**, 595–599 (2006).
30. Rogers, J. A., Someya, T. & Huang, Y. G. Materials and mechanics for stretchable electronics. *Science* **327**, 1603–1607 (2010).

Acknowledgements This work was supported by the National Science Foundation Nano-scale Science and Engineering Center (NSF-NSEC) for Scalable and Integrated Nano Manufacturing (SINAM) (grant no. CMMI-0751621) and by the US Department of Energy, Basic Energy Sciences Energy Frontier Research Center (DoE-LMI-EFRC) under award DOE DE-AC02-05CH11231. M.L. thanks Y. Rao for discussions.

Author Contributions M.L. and X.Z. contributed to the experimental ideas. M.L. fabricated device samples. M.L. and X.Y. carried out measurements, analysed the experimental data and prepared the manuscript. B.G., L.J. and F.W. prepared graphene film. All authors contributed to discussions and manuscript revision.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to X.Z. (xiang@berkeley.edu, for general experimental details) and F.W. (fengwang76@berkeley.edu, for details of graphene synthesis).

Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells

William A. Pastor^{1,2*}, Utz J. Pape^{1,3*}, Yun Huang^{2*}, Hope R. Henderson^{1,2}, Ryan Lister⁴, Myunggon Ko², Erin M. McLoughlin⁵, Yevgeny Brudno⁶, Sahasransu Mahapatra², Philipp Kapranov⁷, Mamta Tahiliani^{1†}, George Q. Daley⁵, X. Shirley Liu³, Joseph R. Ecker⁴, Patrice M. Milos⁷, Suneet Agarwal⁵ & Anjana Rao^{1,2}

5-hydroxymethylcytosine (5hmC) is a modified base present at low levels in diverse cell types in mammals^{1–5}. 5hmC is generated by the TET family of Fe(II) and 2-oxoglutarate-dependent enzymes through oxidation of 5-methylcytosine (5mC)^{1,2,4–7}. 5hmC and TET proteins have been implicated in stem cell biology and cancer^{1,4,5,8,9}, but information on the genome-wide distribution of 5hmC is limited. Here we describe two novel and specific approaches to profile the genomic localization of 5hmC. The first approach, termed GLIB (glucosylation, periodate oxidation, biotinylation) uses a combination of enzymatic and chemical steps to isolate DNA fragments containing as few as a single 5hmC. The second approach involves conversion of 5hmC to cytosine 5-methylenesulphonate (CMS) by treatment of genomic DNA with sodium bisulphite, followed by immunoprecipitation of CMS-containing DNA with a specific anti-serum to CMS⁵. High-throughput sequencing of 5hmC-containing DNA from mouse embryonic stem (ES) cells showed strong enrichment within exons and near transcriptional start sites. 5hmC was especially enriched at the start sites of genes whose promoters bear dual histone 3 lysine 27 trimethylation (H3K27me3) and histone 3 lysine 4 trimethylation (H3K4me3) marks. Our results indicate that 5hmC has a probable role in transcriptional regulation, and suggest a model in which 5hmC contributes to the ‘poised’ chromatin signature found at developmentally-regulated genes in ES cells.

We developed two independent methods for precipitation of 5hmC in genomic DNA. The GLIB method (Fig. 1a) entails addition of a glucose molecule to each 5hmC with T4 phage β -glucosyltransferase (BGT)³ (Supplementary Fig. 1a). The glucose moiety is oxidized with sodium periodate, which converts the vicinal hydroxyl groups to aldehydes¹⁰, and further modified with aldehyde-reactive probe, which adds two biotin molecules to each 5hmC (Fig. 1a). A related strategy, which uses a custom-synthesized UDP-glucose analogue (UDP-6-N3-glucose), was recently used to profile 5hmC distribution in mouse brain¹¹. The second method uses an antibody against cytosine 5-methylenesulphonate (CMS)⁵, produced by reaction of 5hmC with sodium bisulphite (Fig. 1b)¹². Anti-CMS antibodies are more sensitive and less density-dependent than anti-5hmC in DNA dot blot assays⁵. Both methods are specific for DNA containing 5hmC (Supplementary Fig. 1b)⁵.

We examined the ability of GLIB-treated (biotinylated) and bisulphite-treated 5hmC-containing DNA to be pulled down by streptavidin and anti-CMS antisera, respectively. Using varying ratios of dCTP:dhmCTP, we generated 201 base pairs PCR amplicons with differing incorporation of cytosine and 5hmC in identical sequence contexts (Supplementary Table 1). At each dhmCTP:dCTP ratio, the fraction of amplicons that contain no 5hmC, and therefore should not be precipitated, can be calculated using the binomial equation (Supplementary Table 2). Observed and calculated pull-down efficiencies

were very similar (Fig. 1c): even at low densities of 5hmC, more than 90% of DNA fragments calculated to contain a single 5hmC were precipitated after GLIB treatment. Anti-CMS pull-down showed increased density dependence compared to GLIB, but had very low background, such that there was still a strong preference for precipitation of sparsely hydroxymethylated amplicons over unmodified ones (Fig. 1d). The performance of a commercial polyclonal anti-hmC antiserum was inferior to that of anti-CMS, in terms of higher background pull-down of unmodified DNA (3.0% versus 0.06%) as well as greater density dependence (Fig. 1e). By testing PCR amplicons with varying 5mC, we confirmed that the methyl-DNA immunoprecipitation (MeDIP) technique, which uses a monoclonal antibody to 5mC, is extremely density-dependent (Fig. 1f).

We applied the GLIB and anti-CMS techniques to enrich 5hmC-containing regions in genomic DNA using genomic DNA with low, intermediate and high levels of 5hmC (Supplementary Fig. 1c, left panel). For the GLIB and anti-CMS pull-downs, the amount of specifically precipitated genomic DNA was proportional to the relative amount of 5hmC (Supplementary Fig. 1c). The GLIB technique did not produce mutations (Supplementary Table 3), the biotinylated DNA could be efficiently eluted by heating with formamide (Supplementary Fig. 1d), and the biotinylated adduct had a minimal inhibitory effect on PCR at 5–25% hmC density (delay of approximately 0.1 cycles per converted 5hmC residue; Supplementary Fig. 1f). There was no PCR delay with CMS-containing PCR amplicons except at very high CMS levels (Supplementary Fig. 1e), consistent with our previous report that CMS inhibits PCR predominantly at biologically irrelevant sequences where multiple CMS adducts occur in a row¹³.

We investigated the genome-wide localization of 5hmC in murine V6.5 ES cells. For GLIB-treated DNA, we chose Helicos single molecule DNA sequencing, which does not require an amplification step and thus avoids PCR bias^{14,15}. For CMS-enriched genomic DNA, we used an Illumina instrument, as longer read lengths are needed for efficient alignment of bisulphite-treated DNA to the genome¹⁶. With the GLIB method, 119,600 regions of the genome, averaging 1,422 bp in length, showed a substantially higher density of reads in the +BGT as opposed to the –BGT sample; with the CMS method, comparison of enriched to input DNA identified 109,264 enriched regions (average length 1,168 bp). There was high overlap in the enriched regions, here designated 5hmC-enriched regions of the genome (HERGs) (Fig. 1g). Comparing the number of HERGs retrieved by using different fractions of aligned reads yielded a curve that approached an asymptote, suggesting that a majority of hydroxymethylated regions had been identified (Supplementary Fig. 2a).

To determine whether HERGs overlapped with methylated DNA regions, we identified 62,991 5mC-enriched regions of the genome

¹Harvard Medical School, Immune Disease Institute and Program in Cellular and Molecular Medicine, Children’s Hospital Boston, Boston, Massachusetts 02115, USA. ²La Jolla Institute for Allergy & Immunology, La Jolla, California 92037, USA. ³Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, Massachusetts 02115, USA. ⁴Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA. ⁵Division of Hematology/Oncology, Children’s Hospital Boston; Dana-Farber Cancer Institute; Harvard Stem Cell Institute, Boston, Massachusetts 02115, USA. ⁶Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁷Helicos BioSciences Corporation, Cambridge, Massachusetts 02139, USA. [†]Present address: Department of Biochemistry, New York University Langone Medical Centre, New York, New York 10016, USA.

*These authors contributed equally to this work.

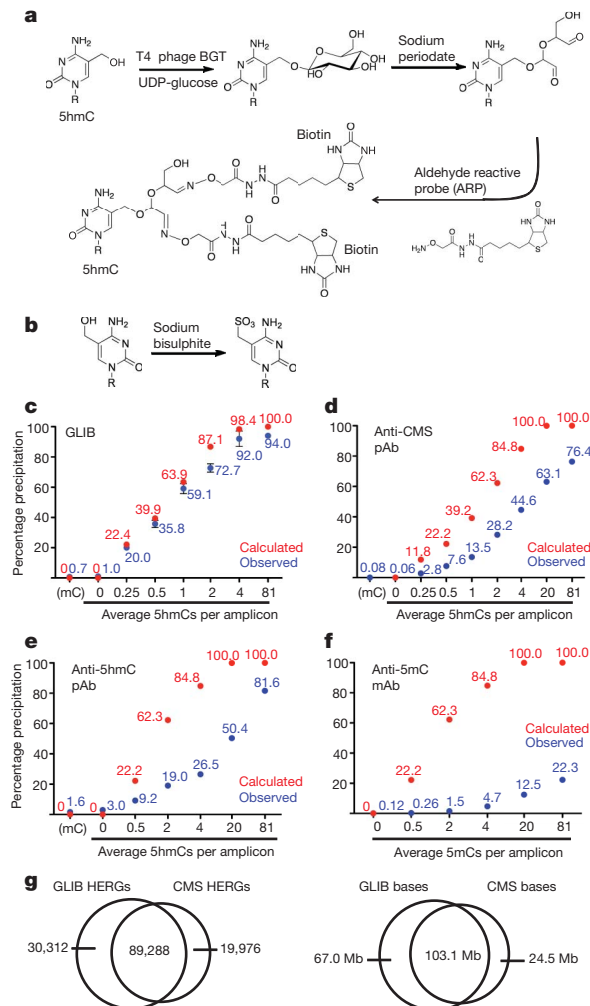


Figure 1 | Comparison of 5hmC enrichment methods. **a**, The GLIB method. Glucose is added to 5hmC by BGT, oxidized with sodium periodate to yield aldehydes, and reacted with the aldehyde reactive probe (ARP), yielding two biotins at the site of every 5hmC. **b**, 5hmC is converted to 5mC by sodium bisulphite. **c–f**, Precipitation of PCR amplicons containing (1) varying amounts of 5hmC by GLIB methodology (**c**), anti-CMS methodology (**d**), or anti-5hmC antibody (**e**); or (2) varying amounts of 5mC by anti-5mC antibody (**f**). pAb, polyclonal antibody; mAb, monoclonal antibody. Between 1 and 6 independent experiments per method, mean percentage input precipitated \pm s.d. is indicated. **g**, Overlap between HERGs identified by the GLIB and anti-CMS methodologies. Left panel, number of HERGs; right panel, number of base pairs contained within HERGs.

(MERGs) by MeDIP. The resulting 5mC profile does not represent a complete map of 5mC in mouse ES cells, but rather is biased towards regions of dense methylation. Statistics pertaining to the GLIB, anti-CMS and MeDIP enrichments are shown in Supplementary Figs 2b–d, the corresponding annotations are provided in Supplementary Tables 4–9, and reads and enrichment for the *Hoxb* locus are provided in Supplementary Table 10. As expected, both HERGs and MERGs contained a high frequency of CG sequences relative to the genome at large (Supplementary Fig. 3a). Intriguingly, HERGs also contained relatively high levels of CAG sequences, the most frequent site of non-CpG methylation in human ES cells¹⁶, and we confirmed that the TET1 catalytic domain is capable of hydroxylating 5mC in CHG and CHH (H = A, T or C) contexts *in vitro* (Supplementary Fig. 3b).

Analysis of the GLIB and anti-CMS HERG sets gave very similar results. We observed a strong correlation between the densities of HERGs and genes on a given chromosome; this trend was less pronounced for MERGs (Fig. 2a). When we compared the distribution of

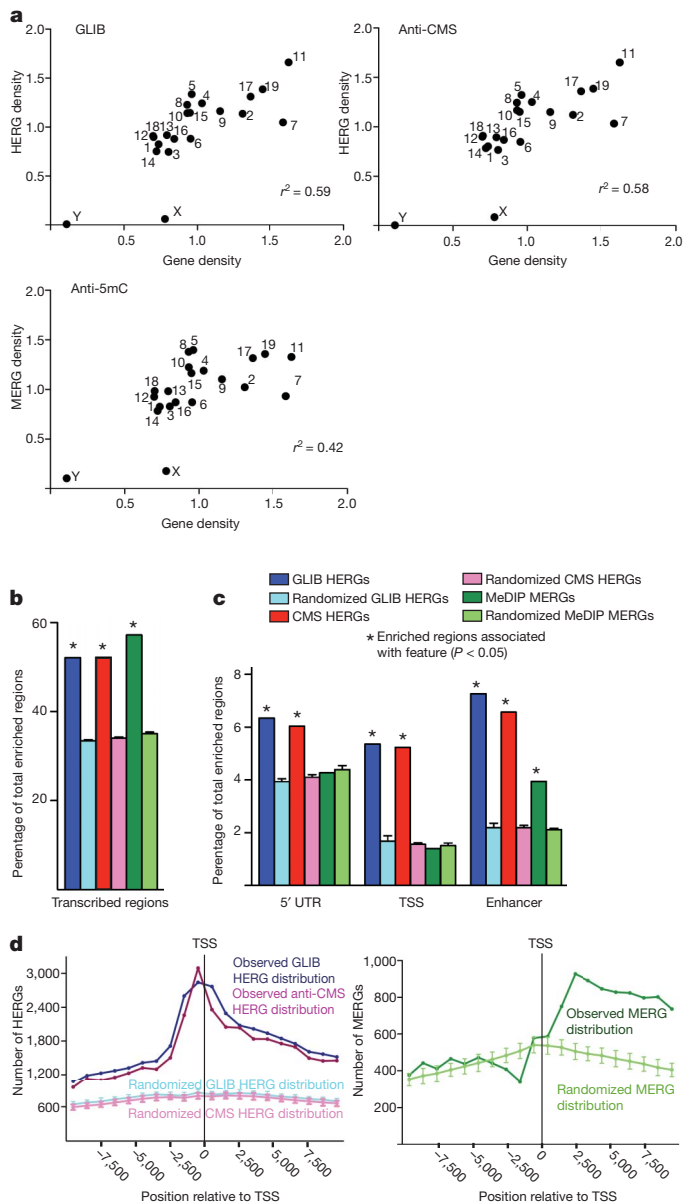


Figure 2 | Genomic distribution of 5hmC or 5mC enriched regions of the genome. **a**, Correlation of HERG or MERG density on each chromosome (y-axis) with gene density in the same chromosome (x-axis). Density is defined as frequency divided by chromosome length. **b**, **c**, Both HERGs and MERGs are enriched in transcribed regions (**b**), whereas HERGs are preferentially enriched at enhancers and the start sites of genes (**c**). The percentage of HERGs or MERGs mapping to the indicated genomic feature (darker bar) is compared with the percentage of randomly chosen sequences mapping to that feature (lighter bar). 5' UTR, 5' untranslated region. TSS, transcription start site (–800 bp to +200 bp relative to start of transcription). See Supplementary Methods for detailed definition of how HERGs or MERGs were classified as mapping to genomic features. **d**, Distribution of HERGs and MERGs relative to the TSS. The centre of each HERG was plotted relative to the nearest TSS in 1,000 bp increments from –10 kb to +10 kb surrounding the TSS.

HERGs and MERGs to the distribution of DNA fragments of equivalent length distributed randomly across the genome, both 5hmC and 5mC were enriched within transcribed regions, particularly exons, which are known to be sites of high CpG density¹⁷ as well as high DNA methylation¹⁸ (Fig. 2b and Supplementary Fig. 3c). However, only 5hmC was enriched at transcription start sites (TSSs) and within the 5' untranslated regions (UTRs) of genes (Fig. 2c). Moreover, 5hmC was relatively more enriched in enhancers (defined by H3K4me1 in the

absence of H3K4me3)¹⁹ than 5mC, strongly indicating a connection between 5hmC and regulatory elements (Fig. 2c). Plotting each HERG as a single point relative to the nearest TSS, we found that 5hmC is heavily enriched both 5' and 3' of the TSS, whereas 5mC is enriched primarily 3' of the TSS (Fig. 2d). These results show a unique distribution of 5hmC in regulatory elements of genes, one that is not explained simply by the distribution of 5mC, the substrate for TET enzymes.

The enrichment of 5hmC at the TSS suggested a role for 5hmC in transcriptional regulation. To evaluate this possibility, we used published data sets on gene expression^{20,21} and histone modification^{22,23} profiles in mouse ES cells to compare the sets of genes with 5hmC or 5mC at their start sites (Supplementary Tables 11–13) to the set of all genes in the genome. 5hmC is preferentially found at promoters with high or intermediate CpG content (Supplementary Fig. 4a), even though high CpG promoters are hypomethylated in ES cells^{16,18,24}. This distribution is consistent with the possibility that TET proteins are preferentially recruited to high CpG regions through their CpG-binding CXXC domains^{6,25}.

In ES cells, genes with 'bivalent' H3K27 and H3K4 trimethylation are transcriptionally inactive but poised for expression upon differentiation to embryoid bodies^{20,26,27}. We found that genes with 5hmC at their start sites were disproportionately likely to contain bivalent domains at their promoters; likewise, a majority (~60%) of genes reported to contain bivalent domains have 5hmC at their start sites (Fig. 3a). 5hmC was less likely to be found at genes with the activating 'H3K4me3 only' mark than is predicted by chance. Moreover, genes with 5hmC at their start sites showed lower expression in murine ES cells than other genes (Fig. 3b) and were more likely to be upregulated upon embryoid body differentiation (Fig. 3c). The correlation of 5hmC with bivalent domains holds even after adjusting for the known relation between promoter CpG content and bivalency²² (Supplementary Fig. 5). Although 5mC at the TSS also correlates with lower gene expression in murine ES cells (Supplementary Fig. 4b), 5mC is not enriched at the promoters of genes with bivalent domains²⁸ (Supplementary Fig. 4c), and genes with high levels of 5mC did not tend to be upregulated upon embryoid body differentiation (Supplementary Fig. 4d). Thus 5hmC is preferentially enriched at the promoters of genes with bivalent histone marks in ES cells, indicating that 5hmC may contribute functionally to the 'poised' but inactive state of these genes in ES cells.

Genes with 5hmC at their start sites are also disproportionately enriched in the set of genes whose promoters bind polycomb repressor complex (PRC) components, and in a majority of genes with the 'H3K27me3' only mark (Fig. 3a). There is a statistically significant correlation between genes that had 5hmC at the TSS and genes that were upregulated upon small interfering RNA-mediated Tet1 depletion⁸ (therefore, negatively regulated by Tet1) (Fig. 3d), indicating that 5hmC in the promoter region has a negative role in the transcription of some genes in ES cells. Unlike 5mC, however, 5hmC is not substantially enriched at sites of heterochromatic H3K9 or H4K20 trimethylation²² (data not shown).

Collectively, our results support a model in which 5hmC and 5mC have different roles in transcription. Like 5mC²⁸, 5hmC at promoters is predictive of lower levels of gene expression. However, 5hmC is uniquely associated with a 'poised' chromatin configuration and with genes that are upregulated upon differentiation, and may thus be involved in priming loci for rapid activation in response to appropriate signals. Activation of lineage-specific genetic loci upon differentiation could occur via a postulated 5mC 'demethylation' pathway (5mC to 5hmC to cytosine)¹ or through recruitment of transcriptional regulators that specifically recognize 5hmC and are activated in response to differentiation signals. The ability to profile 5hmC even at sparsely hydroxymethylated loci will allow a careful evaluation of these possibilities in differentiating cells.

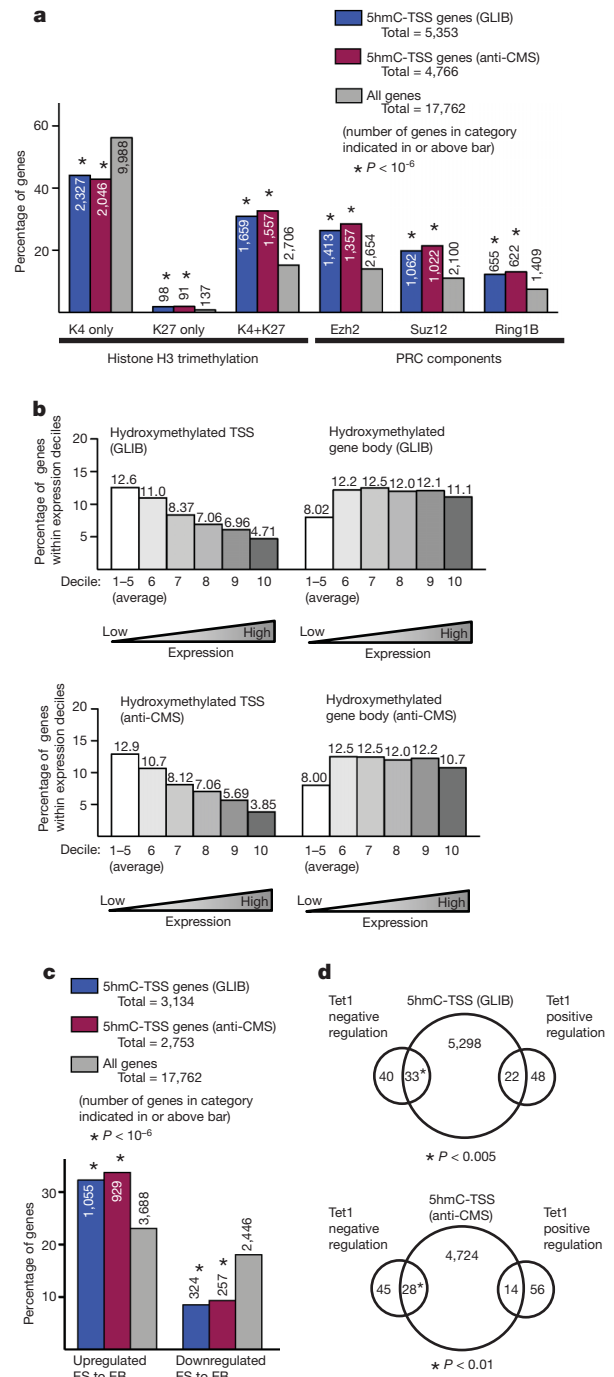


Figure 3 | Properties of HERGs at transcription start sites. **a**, The percentage of genes with 5hmC at the TSS (blue and red bars) reported to contain histone H3 trimethylation (left) or PRC components (right) at their promoters is compared to the fraction of all genes (grey bars) with these promoter marks²². Number of genes in each category is indicated. **b**, HERGs are enriched at the TSSs of genes with low expression in ES cells. All genes were ranked by level of expression in ES cells²¹ and sorted into deciles from lowest to highest. The per cent of genes within the decile category with 5-hmC enriched at the TSS (left) or within gene bodies (right) are shown for each methodology. The first five deciles, which are comprised of genes lacking statistically significant expression, are pooled and averaged in this analysis. **c**, HERGs are enriched at the TSS of genes upregulated upon differentiation to embryoid bodies (EB)²⁶. The percentage of genes with 5-hmC at their TSS (blue bars) that are substantially upregulated or downregulated upon differentiation to EB is compared with the percentage of total genes similarly regulated (grey bars). Number of genes in each category is indicated. **d**, Overlap between genes with 5hmC at the TSS and genes positively or negatively regulated by Tet1 (ref. 8).

METHODS SUMMARY

GLIB precipitation. V6.5 ES cells were lysed and proteins digested by treatment with Proteinase K at 55 °C. DNA was purified by phenol-chloroform extraction and then precipitated with ethanol. RNA was removed with RNase A (Qiagen). Samples were treated with 20 ng BGT per 1 µg DNA at 30 °C for 3 h (50 mM HEPES pH 8.0, 25 mM MgCl₂, 50 µM UDPG for 3 h at 30 °C), then oxidized with 23 mM sodium periodate 16 h at 22 °C in 0.1 M sodium phosphate pH 7.0. Periodate was quenched by the addition of 46 mM sodium sulphite at room temperature for 10 min, then exchanged into 1× PBS and incubated with 2 mM Aldehyde Reactive Probe (Invitrogen) for 1 h at 37 °C. DNA was sequenced with a HeliScope Single Molecule Sequencer. See Supplementary Methods for detailed protocol.

CMS precipitation. The generation of the anti-CMS antibody is described elsewhere⁵. DNA fragments were ligated with methylated adaptors and treated with sodium bisulphite (Qiagen). The DNA was then denatured for 10 min at 95 °C (0.4 M NaOH, 10 mM EDTA), neutralized by addition of cold 2 M ammonium acetate pH 7.0, incubated with anti-CMS antiserum in 1× immunoprecipitation buffer (10 mM sodium phosphate pH 7.0, 140 mM NaCl, 0.05% Triton X-100) for 2 h at 4 °C, and then precipitated with Protein G beads. Precipitated DNA was eluted with Proteinase K, purified by phenol-chloroform extraction, and amplified by 4–6 cycles PCR using Pfu TurboC_x hotstart DNA polymerase (Stratagene). DNA sequencing was carried out using Illumina/Solexa Genome Analyzer II and HiSeq sequencing systems.

Received 4 December 2010; accepted 11 April 2011.

Published online 8 May 2011.

1. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
2. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
3. Szwagierczak, A., Bultmann, S., Schmidt, C. S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **38**, e181 (2010).
4. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
5. Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839–843 (2010).
6. Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8**, 1698–1710 (2009).
7. Loenarz, C. & Schofield, C. J. Oxygenase catalyzed 5-methylcytosine hydroxylation. *Chem. Biol.* **16**, 580–583 (2009).
8. Koh, K. P. *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* **8**, 200–213 (2011).
9. Delhommeau, F. *et al.* Mutation in TET2 in myeloid cancers. *N. Engl. J. Med.* **360**, 2289–2301 (2009).
10. Zhang, H., Li, X. J., Martin, D. B. & Aebersold, R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nature Biotechnol.* **21**, 660–666 (2003).
11. Song, C. X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotechnol.* **29**, 68–72 (2011).
12. Hayatsu, H. & Shiragami, M. Reaction of bisulfite with the 5-hydroxymethyl group in pyrimidines and in phage DNAs. *Biochemistry* **18**, 632–637 (1979).
13. Huang, Y. *et al.* The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE* **5**, e8888 (2010).
14. Harris, T. D. *et al.* Single-molecule DNA sequencing of a viral genome. *Science* **320**, 106–109 (2008).
15. Bowers, J. *et al.* Virtual terminator nucleotides for next-generation DNA sequencing. *Nature Methods* **6**, 593–595 (2009).
16. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
17. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl Acad. Sci. USA* **103**, 1412–1417 (2006).
18. Feng, S. *et al.* Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl Acad. Sci. USA* **107**, 8689–8694 (2010).
19. Creighton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.1016071107 (24 November 2010).
20. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353 (2006).
21. Guttman, M. *et al.* *Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nature Biotechnol.* **28**, 503–510 (2010).
22. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
23. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4**, e1000242 (2008).
24. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770 (2008).
25. Zhang, H. *et al.* TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res.* **20**, 1390–1393 (2010).
26. Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
27. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
28. Fouse, S. D. *et al.* Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**, 160–169 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Ren for assistance in next generation sequencing using the Illumina platform. We thank M. Guttman for making his RNASeq data set available to us. W.A.P. is supported by a predoctoral graduate research fellowship from the National Science Foundation, and Y.H. by a postdoctoral fellowship from the Leukemia and Lymphoma Society. R.L. is supported by a California Institute for Regenerative Medicine Training Grant. This study was supported by the National Institute of Health grants RC1 DA028422, R01 AI44432 and 1 R01 HD065812-01A1 and a grant from the California Institute of Regenerative Medicine (to A.R.), a pilot grant from Harvard Catalyst, The Harvard Clinical and Translational Science Center (NIH Grant 1 UL1 RR 025758-02) and NIH K08 HL089150 (to S.A.), and a grant from the Mary K. Chapman Foundation (to J.R.E.).

Author Contributions W.A.P., Y.B. and S.A. devised the GLIB method. W.A.P., S.A., H.R.H. and E.M.M. optimized the GLIB method. Y.H. generated the anti-CMS antiserum, and Y.H. and W.A.P. optimized the anti-CMS pull-down. W.A.P. and Y.H. grew ES cells. W.A.P. prepared GLIB samples for sequencing, Y.H. prepared CMS samples, H.R.H. performed MedIPs. Helicos sequencing and mapping was performed by P.K. and P.M.M., Illumina sequencing and mapping was performed by R.L. and J.R.E., and U.J.P. was responsible for bioinformatic analysis. M.K. performed the anti-5hmC dot blot. W.A.P. and M.T. performed anti-5hmC pull-downs. H.R.H. and S.M. performed and optimized *in vitro* tests of Tet substrate specificity. W.A.P., S.A. and A.R. wrote the manuscript. S.A. and A.R. coordinated research.

Author Information Data have been deposited at GEO under accession number GSE28682. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.A. (Suneet.Agarwal@childrens.harvard.edu) or A.R. (arao@idi.harvard.edu or arao@iiai.org).

Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors

Pengyu Huang¹, Zhiying He^{1,2}, Shuyi Ji¹, Huawang Sun¹, Dao Xiang², Changcheng Liu^{1,2}, Yiping Hu², Xin Wang^{1,3,4} & Lijian Hui¹

The generation of functional hepatocytes independent of donor liver organs is of great therapeutic interest with regard to regenerative medicine and possible cures for liver disease¹. Induced hepatic differentiation has been achieved previously using embryonic stem cells or induced pluripotent stem cells^{2–8}. Particularly, hepatocytes generated from a patient's own induced pluripotent stem cells could theoretically avoid immunological rejection. However, the induction of hepatocytes from induced pluripotent stem cells is a complicated process that would probably be replaced with the arrival of improved technology. Overexpression of lineage-specific transcription factors directly converts terminally differentiated cells into some other lineages^{9–12}, including neurons¹³, cardiomyocytes¹⁴ and blood progenitors¹⁵; however, it remains unclear whether these lineage-converted cells could repair damaged tissues *in vivo*. Here we demonstrate the direct induction of functional hepatocyte-like (iHep) cells from mouse tail-tip fibroblasts by transduction of Gata4, Hnf1 α and Foxa3, and inactivation of p19^{Arf}. iHep cells show typical epithelial morphology, express hepatic genes and acquire hepatocyte functions. Notably, transplanted iHep cells repopulate the livers of fumarylacetoacetate-hydrolase-deficient (*Fah*^{−/−}) mice and rescue almost half of recipients from death by restoring liver functions. Our study provides a novel strategy to generate functional hepatocyte-like cells for the purpose of liver engineering and regenerative medicine.

Fourteen mouse transcription factors (14TF, Supplementary Table 1) important for liver development and function^{16–19} were transduced into immortalized 3T3 fibroblasts, mouse embryonic fibroblasts (MEFs) and tail-tip fibroblasts (TTFs) via lentiviral infection. The hepatic genes albumin (*Alb*) and *Tdo2* were induced in these cells at day 5 after infection (Supplementary Fig. 1a), indicating that fibroblasts have the potential to be converted to hepatocytes. To ensure that the process is independent of spontaneous immortalization and embryonic progenitors, we focused on TTFs in the following study. Wild-type TTFs showed proliferation arrest and cell death within 7 days after transduction (Supplementary Fig. 1b), thereby inhibiting continuous hepatic conversion. Because p19^{Arf} (also called *Cdkn2a*)-null (p19^{Arf}^{−/−}) hepatocytes proliferate *in vitro* without losing genetic stability²⁰, we used p19^{Arf}^{−/−} TTFs to overcome the proliferative limitation (Fig. 1a). Remarkably, proliferative cells with epithelial morphology were induced from mesenchymal p19^{Arf}^{−/−} TTFs after transduction of 14TF (Supplementary Fig. 1b). Moreover, these cells expressed *Alb*, *Tdo2* and *Ttr* (Supplementary Fig. 1c).

Eleven epithelial colonies, picked up at day 21 after lentiviral transduction, expressed hepatic genes and the exogenous 14TF at different levels (Supplementary Fig. 2). One epithelial colony, ET26, was further characterized (Fig. 1b). ET26 cells expressed hepatic secretory protein genes, cytokeratin genes, epithelial cell adhesion genes and endogenous hepatic transcription factors (Fig. 1c). By contrast, expression of *Col1a1*, *Pdgfrb*, *Postn* and *Fsp1* (also called *S100a4*), genes typical for fibroblast, was downregulated in ET26 cells (Fig. 1c). Functionally, ET26 cells showed glycogen storage as demonstrated by periodic acid-Schiff

(PAS) staining (Fig. 1d) and uptake of DiI-labelled acetylated low-density lipoprotein (DiI-ac-LDL, Fig. 1e). These results indicated that p19^{Arf}^{−/−} TTFs were converted into cells with significant hepatic gene expression and hepatic functions.

Next, we determined the key factors required for hepatic conversion. On the basis of previous reports^{16–19}, we established combinations of six factors (6TF), including Foxa2, Foxa3, Hnf1 α , Hnf4 α , Hnf6 and Gata4, and eight factors (8TF), including 6TF plus Foxa1 and Hlf. Either 6TF or 8TF converted TTFs to epithelial colonies with hepatic gene expression at comparable levels (Supplementary Fig. 3a, b). Upon withdrawal of Hnf6 from 6TF, we found significantly increased

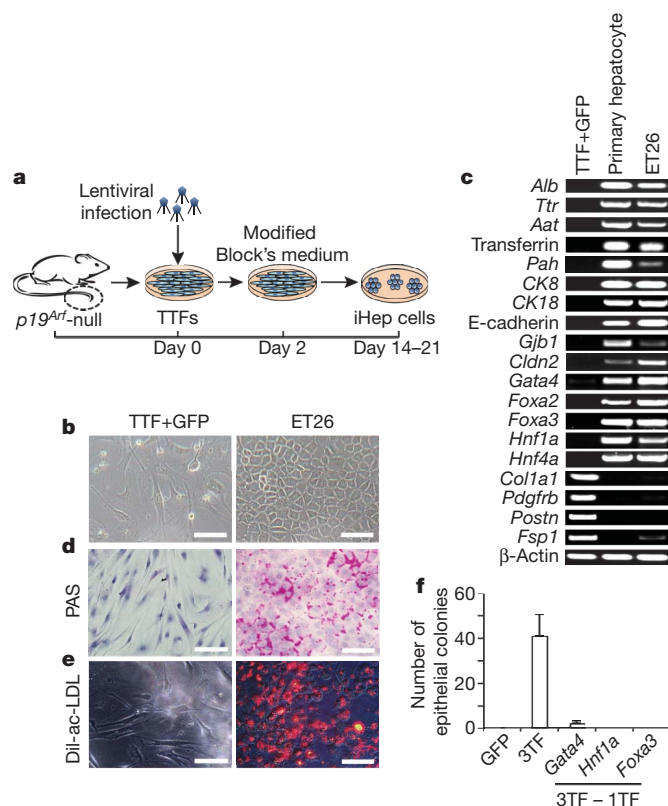


Figure 1 | Three transcription factors induce hepatic conversion of tail-tip fibroblasts. **a**, Experimental design of iHep cell induction. Primary p19^{Arf}^{−/−} TTFs were infected with lentiviruses expressing hepatic transcription factors. Cultures were changed to modified Block's medium 2 days after infection. **b**, Colony ET26 shows epithelial morphology. **c**, Expression of the indicated genes was measured by RT-PCR in ET26 cells, primary hepatocytes and TTFs. **d**, Cytoplasmic accumulation of glycogen was determined by PAS staining (purple cytoplasmic staining). **e**, Intake of DiI-ac-LDL in ET26 cells (red staining). All scale bars: 100 μ m. **f**, Effects of individual factor withdrawal from 3TF on epithelial colony formation. Data are presented as mean + s.d.

¹Laboratory of Molecular Cell Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy for Sciences, Yueyang Road 320, 200031 Shanghai, China.

²Department of Cell Biology, Second Military Medical University, 800 Xiangyin Road, 200433 Shanghai, China. ³Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, 55455 Minnesota, USA. ⁴Stem Cell Institute, University of Minnesota, Minneapolis, 55455 Minnesota, USA.

hepatic gene expression and epithelial colony formation (Supplementary Fig. 3a, b). For the remaining five factors (5TF), removal of *Hnf4 α* further promoted the formation of epithelial colonies (Supplementary Fig. 3c). The remaining four factors were further grouped into two combinations: Gata4, *Hnf1 α* and *Foxa3* (3TF) and Gata4, *Hnf1 α* and *Foxa2* (3TF'). 3TF showed a stronger effect than 3TF' on the induction of hepatic gene expression and epithelial colony formation (Supplementary Fig. 3d and data not shown). Remarkably, 3TF induced endogenous *Foxa2* and *Foxa3* expression (Supplementary Fig. 3d), and removal of any factor from 3TF failed to form epithelial colonies (Fig. 1f). Intriguingly, 3TF triggered *p19^{Arf}*^{-/-} MEFs to express hepatic genes (Supplementary Fig. 4), indicating the potential to induce hepatic conversion of embryonic fibroblasts. Upon RNA-interference-mediated knockdown of *p19^{Arf}*, 3TF also converted wild-type TTFs to epithelial cells with hepatic gene expression (Supplementary Fig. 5).

iHep cells induced by the overexpression of Gata4, *Hnf1 α* and *Foxa3* and the inactivation of *p19^{Arf}* were characterized for their hepatic features. At day 6, the epithelial cells induced by 3TF were positively stained for tight junction protein 1 (Tjp1) and E-cadherin (Fig. 2a–c). At day 14, 23% of epithelial cells were positive for Alb (Supplementary Fig. 6a), indicating an efficient hepatic conversion. The increased expression of hepatic genes over time, for example, *Alb*, *Ttr*, transferrin (*Ttrf*) and *CK18* (also called *Krt18*), showed a progressively enhanced reprogramming (Fig. 2d and Supplementary Fig. 6b, $P < 0.05$). Interestingly, iHep cells also expressed *Afp* and *CK19* (also called *Krt19*) (Fig. 2d). Protein expression of Alb and *Hnf4 α* was confirmed by immunofluorescent staining in iHep cells (Supplementary Fig. 6c, d). Notably, expression levels of exogenous 3TF were markedly decreased during hepatic conversion, indicating that continuous expression of exogenous 3TF is not required (Supplementary Fig. 6e).

Individual iHep colonies showed similar expression patterns of hepatic genes and fibrotic genes (Supplementary Fig. 6f), indicating

a homogeneous conversion among individual TTFs. Although iHep cells expressed *Afp* and *CK19* (Fig. 2d), other hepatoblast marker genes, such as *Lin28b*, *Igf2* and *Dlk1* (ref. 8), were undetectable during hepatic conversion (Supplementary Fig. 7a). Importantly, cytochrome P450 (CYP) enzymes specific to mature hepatocytes were detectable in iHep cells (Supplementary Fig. 7b), suggesting that hepatic conversion undertakes a process without reversion to progenitors. Moreover, iHep cells neither expressed bile duct marker genes nor formed branching bile duct tubes *in vitro* (Supplementary Fig. 7c, d). The marker genes for pancreatic exocrine and endocrine cells and intestinal cells were also undetectable (Supplementary Fig. 7e, f). Therefore, TTFs are not converted to lineages other than hepatocytes.

We compared the global expression profiles among iHep cells, TTFs, MEFs and hepatocytes cultured for 6 days. Pearson correlation analysis showed that iHep cells were clustered with cultured hepatocytes but separated from TTFs and MEFs (Fig. 2e). Microarray data revealed that numerous hepatic functional genes were upregulated in iHep cells compared to TTFs (Supplementary Figs 8 and 9). When compared with cultured hepatocytes, 877 out of 29,153 annotated genes were found to be upregulated in iHep cells, including *Afp*, *CK19*, *Fabp4* and *S100a9*, whereas 817 genes were downregulated, such as *Cyp4b1*, *Cyp2c40* and *Apob* (fold change > 2 , $P < 0.01$, *t*-test). Notably, iHep cells established substantial hepatic functions. iHep cells accumulated PAS-positive glycogen aggregations and transported DiI-ac-LDL into the cytoplasm (Fig. 2f, g). Indocyanine green uptake was found in 20% of iHep cells (Fig. 2h). Furthermore, iHep cells secreted high amounts of Alb into medium (Fig. 2i, $P < 0.05$). Importantly, iHep cells metabolized phenacetin, testosterone and diclofenac (Fig. 2j–l and Supplementary Fig. 10a–c, $P < 0.05$), whereas metabolic activity for bufuralol was undetected (Supplementary Fig. 10d).

Fah^{-/-} mice defective in tyrosine metabolism require 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC) supply for survival^{21–24}. After NTBC withdrawal (NTBC-off), *Fah*^{-/-} mice

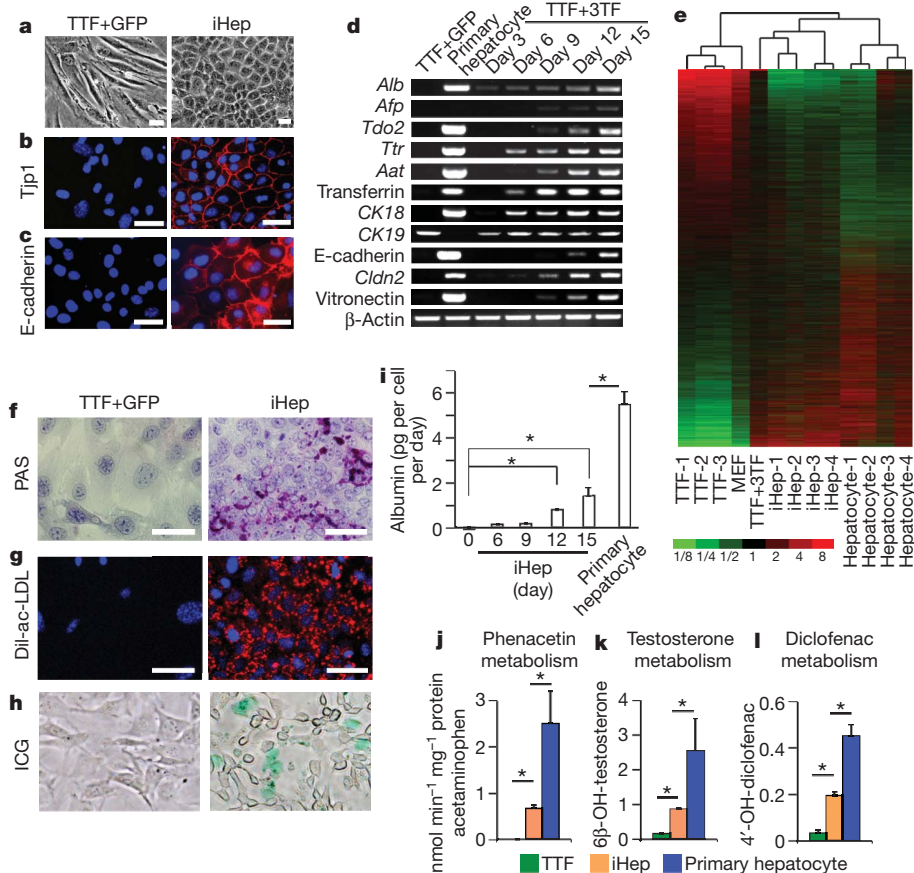


Figure 2 | Characterization of iHep cells *in vitro*.

a, 3TF-induced iHep cells show typical epithelial morphology. **b**, **c**, Epithelial conversion of TTFs was confirmed by immunofluorescent staining of Tjp1 and E-cadherin (red membrane staining). Nuclei are stained blue by DAPI. **d**, Expression of indicated genes was analysed by RT-PCR during the induction of iHep cells. **e**, Global gene expression by cDNA microarray assay. Expression profiles were clustered by a Pearson correlation analysis. Expression levels are depicted in colour. TTF+3TF, 3TF-transduced TTFs without enrichment of epithelial cells. Hepatocyte, hepatocytes cultured for 6 days. **f**, Glycogen storage was assayed by PAS staining. **g**, DiI-ac-LDL uptake in iHep cells. **h**, ICG uptake in iHep cells (green staining). **i**, Secretory albumin protein levels were measured by ELISA during hepatic conversion. **j**–**l**, CYP metabolic activities of iHep cells. The metabolic products of phenacetin (converted to acetaminophen by *Cyp1a2*; **j**), testosterone (converted to 6 β -OH-testosterone by *Cyp3a* enzymes; **k**) and diclofenac (converted to 4'-OH-diclofenac by *Cyp2c* enzymes; **l**) were determined by liquid chromatography-tandem mass spectrometry according to standard curve. *, $P < 0.05$, *t*-test. All scale bars: 50 μ m. Data are presented as mean \pm s.d. in **i**–**l**.

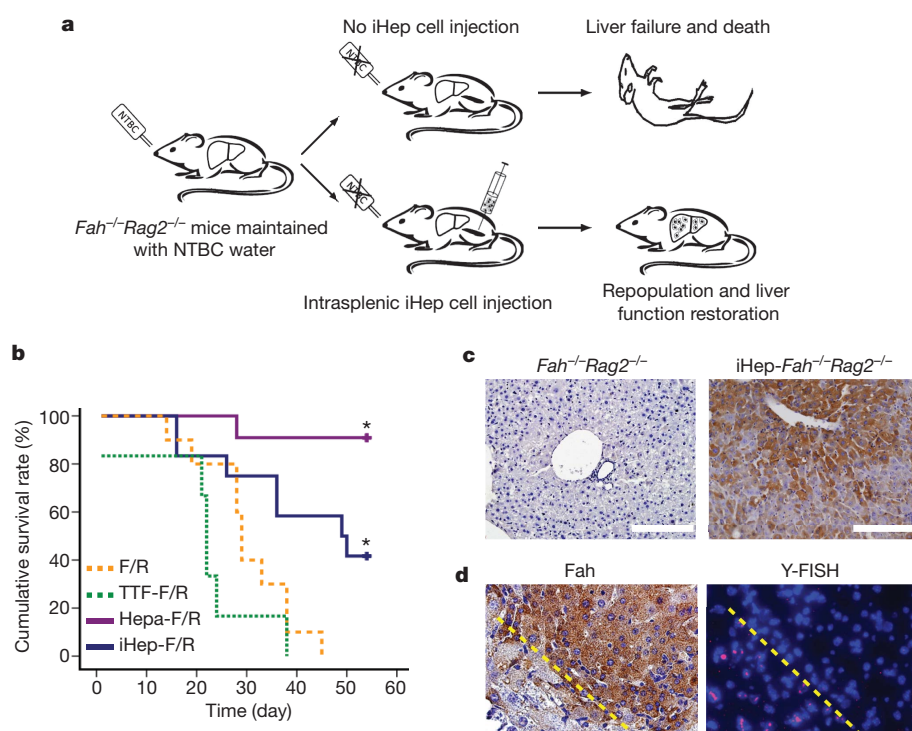


Figure 3 | iHep cell transplantation rescues *Fah*-deficient mice. **a**, Schematic outline of iHep cell transplantation into livers of *Fah*^{-/-}*Rag2*^{-/-} mice.

b, Kaplan-Meier survival curve of primary-hepatocyte-transplanted *Fah*^{-/-}*Rag2*^{-/-} mice (Hepa-F/R, *n* = 10), iHep-cell-transplanted *Fah*^{-/-}*Rag2*^{-/-} mice (iHep-F/R, *n* = 12), TTF-transplanted *Fah*^{-/-}*Rag2*^{-/-} mice (TTF-F/R, *n* = 6) and control *Fah*^{-/-}*Rag2*^{-/-} mice (F/R, *n* = 10) after NTBC withdrawal. *, *P* < 0.02, log-rank test. **c**, Repopulation of iHep cells in *Fah*^{-/-}*Rag2*^{-/-} livers was determined by Fah immunostaining (brown cytoplasmic staining). **d**, Female iHep cells were transplanted into male *Fah*^{-/-}*Rag2*^{-/-} livers. Serial liver sections were stained for both Fah immunostaining and Y-chromosome FISH staining (red dots). The boundary of the *Fah*⁺ nodule is indicated by a dashed yellow line.

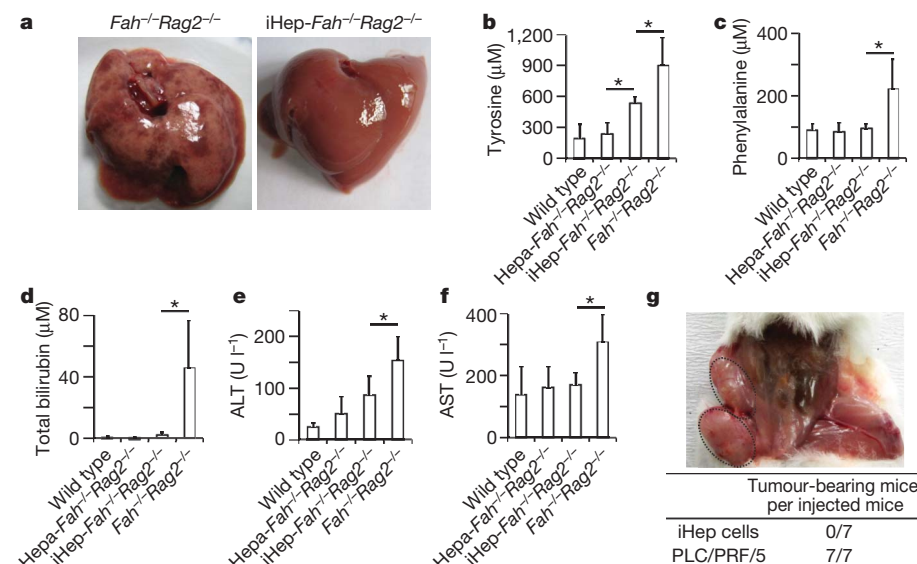
undergo liver failure and death. They can be rescued by transplantation of wild-type primary hepatocytes^{21–24}, representing a useful model to characterize *in vivo* repopulation and functions of iHep cells. Immunodeficient *Fah*^{-/-}*Rag2*^{-/-} mice were used for transplantation to reduce the likelihood of immunological rejection (Fig. 3a, b and Supplementary Fig. 11a). Ten *Fah*^{-/-}*Rag2*^{-/-} mice without transplantation were all dead within 6.5 weeks after NTBC-off and showed continuous loss of body weight (Fig. 3b and Supplementary Fig. 11b). Six *Fah*^{-/-}*Rag2*^{-/-} mice transplanted with *p19*^{Arf}^{-/-} TTFs were also dead after NTBC-off (Fig. 3b). In contrast, 5 out of 12 *Fah*^{-/-}*Rag2*^{-/-} mice transplanted with iHep cells (iHep-*Fah*^{-/-}*Rag2*^{-/-}) were alive 8 weeks after NTBC-off and showed increased body weight (Fig. 3b and Supplementary Fig. 11b, *P* < 0.05). *Fah*-positive (*Fah*⁺) iHep cells engrafting into liver sinusoid comprised between 5% to 80% of total hepatocytes in iHep-*Fah*^{-/-}*Rag2*^{-/-} livers (Fig. 3c and Supplementary Fig. 11c). Moreover, *Fah*-wild-type and *p19*^{Arf}-null alleles were detected in iHep-*Fah*^{-/-}*Rag2*^{-/-} livers by genomic PCR (Supplementary Fig. 11d). To exclude the possibility of cell fusion between iHep and host

cells, we stained the Y chromosome in male livers transplanted with female iHep cells. Twenty-five *Fah*⁺ nodules in four male recipients were characterized and were all negative for Y-chromosome staining, confirming that iHep cells do not fuse with host cells (Fig. 3d and Supplementary Fig. 11e). These results indicate that transplanted iHep cells can repopulate and rescue *Fah*^{-/-}*Rag2*^{-/-} recipients.

Macroscopically, iHep-*Fah*^{-/-}*Rag2*^{-/-} livers are normal and healthy, whereas livers from NTBC-off *Fah*^{-/-}*Rag2*^{-/-} control mice were swelled with many necrotic lesions (Fig. 4a). The hexagonal hepatic lobule was destructed due to massive cell death in NTBC-off *Fah*^{-/-}*Rag2*^{-/-} livers (Supplementary Fig. 12a). In contrast, iHep cell repopulation restored liver architecture without apparent cell death (Supplementary Fig. 12a, b). Remarkably, both repopulated iHep cells and repopulated primary hepatocytes expressed Alb and other hepatic genes at comparable levels in *Fah*^{-/-}*Rag2*^{-/-} recipient mice (Supplementary Fig. 12c, d). Moreover, serum levels of tyrosine, phenylalanine, ornithine, alanine and glycine were markedly reduced in iHep-*Fah*^{-/-}*Rag2*^{-/-} mice compared to NTBC-off *Fah*^{-/-}*Rag2*^{-/-}

Figure 4 | iHep cells restore liver functions of *Fah*^{-/-}*Rag2*^{-/-} mice.

a, Representative pictures of whole livers from *Fah*^{-/-}*Rag2*^{-/-} and iHep-*Fah*^{-/-}*Rag2*^{-/-} mice. **b–f**, Serum levels of tyrosine (b), phenylalanine (c), total bilirubin (d), ALT (e) and AST (f) in wild-type (*n* = 6), Hepa-*Fah*^{-/-}*Rag2*^{-/-} (*n* = 5), iHep-*Fah*^{-/-}*Rag2*^{-/-} (*n* = 5, sera collected 8 weeks after iHep transplantation) and control *Fah*^{-/-}*Rag2*^{-/-} mice (*n* = 4, sera collected upon losing 20% of body weight). *, *P* < 0.05, *t*-test. Data are presented as mean ± s.d. **g**, iHep and PLC/PRF/5 cells (human hepatoma cell line) were subcutaneously transplanted into the left and right flanks of NOD/SCID mice, respectively. PLC/PRF/5-generated tumours are indicated by the dotted ovals.



mice (Fig. 4b, c, Supplementary Fig. 12e–g, and Supplementary Table 2, $P < 0.05$). iHep-*Fah*^{-/-}*Rag2*^{-/-} mice also showed decreased levels of total bilirubin, alanine aminotransferase (ALT) and aspartate aminotransferase (AST) (Fig. 4d–f and Supplementary Table 3, $P < 0.05$). These data demonstrate that iHep cell transplantation substantially improves liver functions of NTBC-off *Fah*^{-/-}*Rag2*^{-/-} mice.

Tumours were not found in iHep-*Fah*^{-/-}*Rag2*^{-/-} livers 2 months after transplantation. Ki67 staining revealed that iHep cells ceased proliferation 8 weeks after transplantation (Supplementary Fig. 13a). Moreover, iHep cells did not form tumours 8 weeks after subcutaneous xenograft in NOD/SCID mice (Fig. 4g). A total of 20 out of 25 analysed iHep cells displayed 40 chromosomes after 17 passages (Supplementary Fig. 13b), which was comparable with results from wild-type cells (data not shown). These results indicate that iHep cells do not seem to be tumour prone.

To our knowledge, this is the first time that adult fibroblasts have been directly converted to functional iHep cells. Together with previous findings^{9–15}, our results prove the general principle that cell lineages can be converted by regulating the transcriptional network. We identified the combination of Gata4, Foxa3 and Hnf1 α as being sufficient to induce hepatic conversion. Gata4 and Foxa3 probably act as 'pioneer factors' to trigger a global chromatin modification during hepatic conversion^{25,26}. Hnf1 α probably stabilizes the hepatic gene expression, as Hnf1 α , Foxa2 and Hnf4 α occupy each other's promoters and maintain the hepatic phenotype^{16,27}. Moreover, we obtained proliferative iHep cells under the condition of inactivating p19^{Arf}. Because p19^{Arf} is a key component of the cellular senescence pathway that inhibits induced pluripotent stem cell reprogramming²⁸, it would be of interest to characterize whether inactivating other components of this pathway, such as p38 α ²⁹, could be used to facilitate hepatic conversion.

iHep cells showed an expression profile and hepatic function close to those of mature hepatocytes. Interestingly, some CYP genes were not induced in iHep cells, and *CK19* and *Afp* were upregulated. Moreover, iHep cell transplantation showed that the rescue was partial, suggesting that iHep cells are not identical to hepatocytes. Nevertheless, in contrast with any other cell-type conversion via lineage-specific transcription factors^{13–15}, the *in vivo* function of iHep cells has been rigorously proven. More importantly, iHep cells do not seem to be prone to tumour formation. Thus, iHep cells represent an alternative source of hepatocytes for disease modelling, transplantation and tissue engineering. To apply this approach for the purpose of regenerative medicine, future studies will need to address whether human fibroblasts and other cell types could be successfully converted to functional iHep cells.

METHODS SUMMARY

p19^{Arf}^{-/-}, *Fah*^{-/-}*Rag2*^{-/-} and NOD/SCID mice were maintained according to institutional regulation. *p19*^{Arf}^{-/-} TTFs between passage 7 and 9 were used for induction of iHep cells. *p19*^{Arf}^{-/-} TTFs were seeded on collagen-I-coated dishes and infected with lentiviruses expressing transcription factors. Cells were then cultured in Block's medium containing 0.1 μ M dexamethasone, 20 μ g l⁻¹ TGF- α , 10 μ g l⁻¹ EGF, 4.2 mg l⁻¹ insulin, 3.8 mg l⁻¹ human transferrin, and 5 μ g l⁻¹ sodium selenite. Fourteen days after infection, we treated cells with 0.01% trypsin and discarded detached fibroblastic cells to enrich the epithelial cells. iHep cells were transplanted into spleens of *Fah*^{-/-}*Rag2*^{-/-} mice at the age of 8–12 weeks. We intrasplenically injected 8.33×10^5 iHep cells into *Fah*^{-/-}*Rag2*^{-/-} mice. NTBC was withdrawn from the drinking water after transplantation. Surviving *Fah*^{-/-}*Rag2*^{-/-} mice transplanted with iHep cells were killed 8 weeks after the surgery to collect peripheral blood and liver samples. All animal experiments were performed according to institutional regulations. Microarray data have been deposited in the Gene Expression Omnibus database (GSE23635).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 August 2010; accepted 15 April 2011.

Published online 11 May 2011.

1. Zern, M. A. Cell transplantation to replace whole liver transplantation. *Gastroenterology* **136**, 767–769 (2009).
2. Gouon-Evans, V. et al. BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm. *Nature Biotechnol.* **24**, 1402–1411 (2006).

3. Soto-Gutiérrez, A. et al. Reversal of mouse hepatic failure using an implanted liver-assist device containing ES cell-derived hepatocytes. *Nature Biotechnol.* **24**, 1412–1419 (2006).
4. Duan, Y. et al. Differentiation and enrichment of hepatocyte-like cells from human embryonic stem cells *in vitro* and *in vivo*. *Stem Cells* **25**, 3058–3068 (2007).
5. Basma, H. et al. Differentiation and transplantation of human embryonic stem cell-derived hepatocytes. *Gastroenterology* **136**, 990–999 (2009).
6. Si-Tayeb, K. et al. Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology* **51**, 297–305 (2010).
7. Sullivan, G. J. et al. Generation of functional human hepatic endoderm from human induced pluripotent stem cells. *Hepatology* **51**, 329–335 (2010).
8. Li, F. et al. Hepatoblast-like progenitor cells derived from embryonic stem cells can repopulate livers of mice. *Gastroenterology* **139**, 2158–2169 (2010).
9. Shen, C. N., Slack, J. M. & Tosh, D. Molecular basis of transdifferentiation of pancreas to liver. *Nature Cell Biol.* **2**, 879–887 (2000).
10. Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663–676 (2004).
11. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. *In vivo* reprogramming of adult pancreatic exocrine cells to β -cells. *Nature* **455**, 627–632 (2008).
12. Feng, R. et al. PU.1 and C/EBP α / β convert fibroblasts into macrophage-like cells. *Proc. Natl Acad. Sci. USA* **105**, 6057–6062 (2008).
13. Vierbuchen, T. et al. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
14. Ieda, M. et al. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386 (2010).
15. Szabo, E. et al. Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* **468**, 521–526 (2010).
16. Kyrnizi, I. et al. Plasticity and expanding complexity of the hepatic transcription factor network during liver development. *Genes Dev.* **20**, 2293–2305 (2006).
17. Zaret, K. S. Genetic programming of liver and pancreas progenitors: lessons for stem-cell differentiation. *Nature Rev. Genet.* **9**, 329–340 (2008).
18. Schrem, H., Klempnauer, J. & Borlak, J. Liver-enriched transcription factors in liver function and development. Part I: the hepatocyte nuclear factor network and liver-specific gene expression. *Pharmacol. Rev.* **54**, 129–158 (2002).
19. Schrem, H., Klempnauer, J. & Borlak, J. Liver-enriched transcription factors in liver function and development. Part II: the C/EBPs and D site-binding protein in cell cycle control, carcinogenesis, circadian gene regulation, liver regeneration, apoptosis, and liver-specific gene regulation. *Pharmacol. Rev.* **56**, 291–330 (2004).
20. Mikula, M. et al. Immortalized p19ARF null hepatocytes restore liver injury and generate hepatic progenitors after transplantation. *Hepatology* **39**, 628–634 (2004).
21. Grompe, M. et al. Loss of fumarylacetoacetate hydrolase is responsible for the neonatal hepatic dysfunction phenotype of lethal albino mice. *Genes Dev.* **7** (12A), 2298–2307 (1993).
22. Wang, X. et al. The origin and liver repopulating capacity of murine oval cells. *Proc. Natl Acad. Sci. USA* **100** (Suppl. 1), 11881–11888 (2003).
23. Grompe, M. et al. Pharmacological correction of neonatal lethal hepatic dysfunction in a murine model of hereditary tyrosinaemia type I. *Nature Genet.* **10**, 453–460 (1995).
24. Overturf, K. et al. Hepatocytes corrected by gene therapy are selected *in vivo* in a murine model of hereditary tyrosinaemia type I. *Nature Genet.* **12**, 266–273 (1996).
25. Zaret, K. S. et al. Pioneer factors, genetic competence, and inductive signaling: programming liver and pancreas progenitors from the endoderm. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 119–126 (2008).
26. Cirillo, L. A. et al. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
27. Odom, D. T. et al. Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303**, 1378–1381 (2004).
28. Li, H. et al. The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* **460**, 1136–1139 (2009).
29. Hui, L. et al. p38 α suppresses normal and cancer cell proliferation by antagonizing the JNK-c-Jun pathway. *Nature Genet.* **39**, 741–749 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank X. Chen, J. Cen, L. Zhang and J. Yuan for technical support, and F. Tang and D. Li for comments. The laboratory of L.H. is funded by the National Science Foundation of China (91019014) and the Chinese Academy of Sciences (KSCX2-YW-R-241, XDA01010402 and the Hundred Talents Program). The laboratory of X.W. is supported by the National Science Foundation of China (30801115), Chinese National 863 Plan Project (2006AA02Z474) and National Key Basic Research and Development Program of China (2007CB947102 and 2009CB941100).

Author Contributions L.H. conceived the project. P.H. performed most of the experiments. S.J. analysed the *in vitro* functions of iHep cells. H.S. analysed gene expression of iHep cells. L.H., X.W., P.H. and Z.H. designed the experiments for characterizing *in vivo* functions of iHep cells. P.H., Z.H., D.X., C.L. and Y.H. performed the *in vivo* experiments. L.H. and P.H. analysed the data. L.H., P.H. and X.W. wrote the manuscript.

Author Information Microarray data have been deposited in the Gene Expression Omnibus database under accession number GSE23635. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.H. (ljhui@sibs.ac.cn).

METHODS

Mice. $p19^{Arf-/-}$ mice, $Fah^{-/-}Rag2^{-/-}$ mice and NOD/SCID mice were maintained in specific pathogen-free husbandry. $Fah^{-/-}Rag2^{-/-}$ mice were fed with drinking water containing 7.5 mg l^{-1} NTBC. The genetic background for $p19^{Arf-/-}$ and $Fah^{-/-}Rag2^{-/-}$ mice was C57Bl6/J \times 129Sv. $Fah^{-/-}Rag2^{-/-}$ mice were used as the recipient to reduce immunological rejection of iHep cells after transplantation.

Molecular cloning and lentivirus production. A multi-cloning site (CGGGA TCCCGCGCGCCGACTAGTCGACGCGTCGAGGTAACCTACGACCGGT TT) was inserted into the PmeI restriction site of lentiviral vector pWPI (obtained from Addgene). cDNAs of candidate genes were cloned into the modified pWPI plasmid. For $p19^{Arf}$ shRNA expression, DNA oligonucleotides encoding $p19^{Arf}$ shRNA (CCGGGTGAACATGTTGTTGAGGCTAGGATCCTAGCCTCAACA ACATGTTTACATTTTGG) were inserted into the AgeI and EcoRI restriction sites of the pLKO.1 plasmid. Constructed pWPI or pLKO.1 plasmids were then introduced to 293FT cells together with packaging plasmid psPAX2 (Addgene) and envelope plasmid pMD2.G (Addgene). After 48 h incubation, the medium containing lentiviruses was collected and passed through a $0.45\text{ }\mu\text{m}$ filter.

Fibroblast culture and bile duct induction. To isolate tail-tip fibroblasts, 5 cm of tail were cut from two-month-old mice. We peeled the dermis and minced tails into 1-cm pieces. Two pieces were placed per 60-mm collagen-I-coated dish in 5 ml DMEM (Sigma-Aldrich) containing 10% FBS (Sigma-Aldrich). After 5 days incubation, fibroblasts that migrated out of the tails were transferred to new collagen-I-coated dishes. We used TTFs between passage 7 and 9 for iHep cell induction. Embryonic fibroblasts were isolated from E13.5 embryos. Head and visceral tissue were dissected and removed. The remaining tissues were minced and incubated with 0.25% trypsin (Gibco) at 37°C for 15 min. Isolated cells were plated onto a 60-mm collagen-I-coated dish in 5 ml DMEM containing 10% FBS. We used MEFs at passage 3 for lentiviral infection.

For bile duct differentiation, 1×10^4 cells were re-suspended in 1 ml DMEM/F12 medium with 1 ml freshly prepared collagen gel solution and poured into a 35-mm dish. After gel solidification, cells were cultured with 1.5 ml DMEM/F12 supplemented with 10% FBS, $1 \times \text{ITS}$, 20 ng ml^{-1} HGF for 3 days.

Primary hepatocyte isolation and culture. Adult mice were subjected to standard two-step collagenase perfusion for isolation of primary hepatocytes. Briefly, the liver was pre-perfused through the portal vein with calcium-free buffer (0.5 mM EGTA, $1 \times \text{EBSS}$ without Ca^{2+} and Mg^{2+}) and then perfused with collagenase (0.2 mg ml^{-1} collagenase type IV (Sigma), 10 mM HEPES, $1 \times \text{EBSS}$ with Ca^{2+} and Mg^{2+}). Parenchymal cells were purified by Percoll buffer (90% Percoll (Sigma), $1 \times \text{EBSS}$) at low-speed centrifugation ($1,500\text{ r.p.m.}$, 10 min). Viability of isolated hepatocytes was around 90% as determined by Trypan blue. For microarray analysis, $p19^{Arf-/-}$ primary hepatocytes were cultured in modified Block's medium supplemented with $0.1\text{ }\mu\text{M}$ dexamethasone, $20\text{ }\mu\text{g l}^{-1}$ TGF- α , $10\text{ }\mu\text{g l}^{-1}$ EGF, 4.2 mg l^{-1} insulin, 3.8 mg l^{-1} human transferrin and $5\text{ }\mu\text{g l}^{-1}$ sodium selenite in collagen-I-coated dishes for 6 days before harvesting for RNA extraction. For other experiments, $p19^{Arf-/-}$ primary hepatocytes were immediately lysated in Trizol for total RNA isolation.

PCR. For most experiments, total RNA was isolated from cells by Trizol (Invitrogen). For RNA extraction from formalin-fixed-paraffin-embedded (FFPE) tissues, four serial sections mounted on polyethylene terephthalate (PET) membrane frame slides were deparaffinized and air dried. The first section was stained with anti-Fah antibody to identify the repopulated Fah^+ nodules. On the basis of the result of Fah immunostaining in the first section, Fah^+ tissues within the nodules were microdissected from the following three sections by a Leica LMD7000 Laser Microdissection Microscope (Leica Microsystems) with laser intensity of 45 and speed of 5. After microdissection, the remaining sections on the slides were further stained with anti-Fah antibody to confirm that only tissues inside Fah^+ nodules were separated. Microdissected tissues from the same Fah^+ nodule were pooled together for total RNA extraction using RNeasy FFPE Kit (Qiagen).

A total of $1\text{ }\mu\text{g}$ RNA was reverse transcribed into cDNA with M-MLV Reverse Transcriptase (Promega) according to the manufacturer's instructions. For DNA extraction from formalin-fixed-paraffin-embedded tissues, the QIAamp DNA FFPE Tissue Kit (Qiagen) was applied according to the manufacturer's instructions. PCR was performed with HiFi Taq polymerase (TransGen). Quantitative real-time PCR was performed with SYBR Premix Ex Taq (TaKaRa) on an ABI 7500 fast real-time PCR system (Applied Biosystems). Primer sequences will be provided upon request.

Immunofluorescence. For immunofluorescence staining, the cells were fixed with 4% paraformaldehyde for 15 min at room temperature, and then incubated with PBS containing 0.2% Triton X-100 (Sigma) for 15 min. Cells were then washed three times with PBS. After being blocked by 3% BSA in PBS for 60 min at room temperature, cells were incubated with primary antibodies at 4°C

overnight, washed three times with PBS, and then incubated with appropriate fluorescence-conjugated secondary antibody for 60 min at room temperature in the dark. Nuclei were stained with DAPI (Sigma). Primary and secondary antibodies were diluted in PBS containing 3% BSA. Antibodies used for immunofluorescence are as follows: mouse anti-Tjp1 (Invitrogen, 1:750), rabbit anti-E-cadherin (Cell Signaling, 1:500), mouse anti-albumin (R&D, 1:200), goat anti-Hnf4 α (Santa Cruz, 1:200). Cy5-conjugated goat anti-mouse IgG (Jackson Laboratories, 1:1,000), Cy3-conjugated goat anti-rabbit IgG (Jackson Laboratories, 1:1,000), Cy3-conjugated donkey anti-goat IgG (Jackson Lab, 1:1,000). For Y-chromosome fluorescent *in situ* hybridization (FISH), liver samples of male $Fah^{-/-}Rag2^{-/-}$ mice transplanted with female iHep cells were embedded in paraffin and hybridized with mouse Y-chromosome probe (ID Labs Inc., Canada) according to the manufacturer's instruction.

FACS analyses. For intracellular staining of albumin, 10^6 cells were harvested and fixed with 4% PFA for 30 min, and then permeabilized in staining buffer (PBS with 10% FBS and 0.5% saponin) for 10 min. Cells were then incubated with primary antibody (anti-albumin, R&D) for 30 min in staining buffer, followed with secondary antibody (Cy5-conjugated goat anti-mouse IgG, Jackson Laboratories) incubation for 30 min. Cells were analysed by the Calibur flow cytometer (Becton Dickinson). Data were analysed with Windows Multiple Document Interface for Flow Cytometry (WinMDI, version 2.9).

PAS stain, DiI-ac-LDL and ICG uptake assays, Alb ELISA and CYP metabolism assay. Cells were stained by periodic acid-Schiff (PAS, Sigma) and DiI-ac-LDL (Invitrogen) following the manufacturer's instructions. For the indocyanine green (ICG, Sigma) uptake assay, cells were cultured in the medium supplemented with progesterone, pregnenolone-16 α -carbonitrile and 8-bromo cAMP for 2 days. Cells had their medium changed with 1 mg ml^{-1} ICG and were incubated at 37°C for 1 h, followed by washing with PBS three times.

To determine Alb secretion, TTFs transduced with three factors were cultured in the medium without phenol red. Culture supernatant was collected 24 h after medium change. The amount of Alb in the supernatant was determined by the mouse albumin ELISA kit (Bethyl Laboratory) according to the manufacturer's instructions.

For the measurement of CYP enzyme activities, TTFs and iHep cells were cultured in the medium with $50\text{ }\mu\text{M}$ 3-methylcholanthrene for 48 h. Cells were dissociated and incubated with substrate in $200\text{ }\mu\text{l}$ incubation medium at different concentrations for 3 h at 37°C . To stop the reaction, $800\text{ }\mu\text{l}$ cold methanol was added and centrifuged. The supernatants were collected for measurement of indicated productions by LC-MS/MS (Agilent 1200 HPLC and ABI 4000 mass-spectrometer). Freshly isolated hepatocytes were used as a positive control. Total cell protein amount was used to normalize the data. Substrates and metabolic products for standard were commercially purchased: phenacetin, diclofenac, bufuralol, acetaminophen, 4'-OH diclofenac (Sigma), testosterone (Fluka), 6 β -OH-testosterone (Cerilliant) and 1'-OH-bufuralol (Toronto Research Chemicals).

Microarray analysis. Total RNA extracted from $p19^{Arf-/-}$ TTFs, $p19^{Arf-/-}$ MEFs, $p19^{Arf-/-}$ hepatocytes cultured for 6 days, 3TF-transduced $p19^{Arf-/-}$ TTFs without enrichment of epithelial cells, and iHep cells from different experiments was hybridized to whole mouse gene expression microarray (Agilent) under the manufacturer's instruction. Data were normalized by Gene-Spring (Agilent). Microarray hybridization and analysis were carried out by ShanghaiBio Cooperation. Out of 29,153 annotated genes, 11,797 genes for which expression levels were at least twofold different between $p19^{Arf-/-}$ TTFs and primary $p19^{Arf-/-}$ hepatocytes were selected for analyses. Hierarchical clustering of samples was performed by Cluster 3.0 software. Average linkage with the uncentred correlation similarity metric was used for the clustering of samples. Original data were uploaded to the Gene Expression Omnibus database (accession number GSE23635).

In vivo function analysis. $Fah^{-/-}Rag2^{-/-}$ mice were maintained with 7.5 mg l^{-1} NTBC in the drinking water. According to our previous experience with primary hepatocyte and ES-cell-derived hepatoblast transplantation, 8.33×10^5 iHep cells and 8.33×10^5 $p19^{Arf-/-}$ TTFs were transplanted into the spleens of $Fah^{-/-}Rag2^{-/-}$ mice at the age of 8–12 weeks, respectively. NTBC was withdrawn from the drinking water after cell transplantation. Ten $Fah^{-/-}Rag2^{-/-}$ mice without any transplantation also had NTBC withdrawn as a control. A survival curve was generated by SPSS for windows using Kaplan-Meier method. Eight weeks after transplantation, the blood of surviving iHep-cell-transplanted $Fah^{-/-}Rag2^{-/-}$ mice was collected from the retro-orbital sinus and centrifuged at $12,000\text{ r.p.m.}$ for 15 min. The serum was frozen at -80°C until biochemical analyses. Total bilirubin, albumin, ALT, AST, blood urea nitrogen and creatinine were measured by 7600-020 clinical analyser (Hitachi). Amino acids were quantified by liquid chromatography-mass spectrometry ABI 3200 Q TRAP LC-MS/MS system (Applied Biosystem). After blood collection, mice were killed by cervical dislocation and livers were harvested, fixed and stained with Fah polyclonal antibody or haematoxylin and eosin as previously described. Blood and liver samples of control NTBC-off $Fah^{-/-}Rag2^{-/-}$ mice were collected after losing 20% body weight.

Tumour generation assay. The human hepatoma cell line PLC/PRF/5 was cultured in the same medium as iHep cells. iHep cells were induced and enriched as described above. After 21 days induction, cells were detached by trypsin and suspended in PBS. Seven NOD/SCID mice respectively were injected with 5×10^6 iHep cells in the left subcutaneous flank and 5×10^6 PLC/PRF/5 cells in the right subcutaneous flank. Tumour numbers were counted 8 weeks after injection.

Statistics. All data are presented as mean + s.d. For most statistical evaluation, an unpaired Student's *t*-test was applied for calculating statistical probability in this study. For survival analysis, the Mantel–Cox log-rank test was applied. Statistical calculation was performed using Statistical Program for Social Sciences software (SPSS, IBM). For all statistics, data from at least three independent samples or repeated experiments were used.